



UNIVERSITÀ DI PISA

**DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA DEI SISTEMI,
DEL TERRITORIO E DELLE COSTRUZIONI**

**RELAZIONE PER IL CONSEGUIMENTO DELLA
LAUREA MAGISTRALE IN INGEGNERIA GESTIONALE**

***Controllo Statistico Multivariato di Processo:
Applicazione ad un Caso Reale***

RELATORI

Prof. Ing. Riccardo Dulmin

Dipartimento di Ingegneria dell'Energia

dei Sistemi, del Territorio e delle Costruzioni

Marco Calderisi Phd

IL CANDIDATO

Sara Scialanca

sara.scialanca@gmail.com

Sessione di Laurea del 25/11/2015
Anno Accademico 2014/2015
Consultazione *NON* consentita

Sommario

Questo lavoro di Tesi è stato realizzato grazie ad un tirocinio di sei mesi tenuto presso l'azienda Kode srl, con sede a Pisa. L' Obiettivo della Tesi è fornire un'applicazione della tecnica di Controllo Statistico Multivariato di Processo, al fine di sviluppare un modello matematico in grado di descrivere le relazioni tra le variabili di processo e le proprietà dei lotti in produzione in riferimento ai requisiti di specifica. La metodologia di monitoraggio sviluppata è stata pensata per essere eseguita online tra un lotto e il successivo. Il caso di studio che è stato analizzato, è un Processo di Miscelazione di tipo batch per la produzione di Pellets in materiale plastico. I dati di processo sono stati forniti dalla Mixer Spa, azienda produttrice di Pellets di materiale plastico. Lo strumento di monitoraggio, costruito utilizzando il linguaggio di programmazione statistica R, è inserito all'interno di un progetto il cui scopo è la creazione di un sistema di controllo di processo da applicare all'intera linea produttiva dell'azienda Mixer.

Abstract

This thesis work was made possible thanks to a six-month internship at the company held Kode Ltd., based in Pisa. The objective of the thesis is to provide an application of the technique of Multivariate Statistical Process Control, in order to develop a mathematical model able to describe the relationships between process variables and properties of batch production in reference to the specification requirements. The monitoring methodology developed is designed to be performed online from one batch to the next. The case study which has been analyzed, is a process of mixing of batch type for the production of pellets of plastic material. The process data were provided by Mixer Spa, a manufacturer of plastic pellets. The monitoring tool, built using the R , which is integrated within a project whose aim is to create a system of process control applied to the whole production line of the company Mixer.

Indice Generale

Introduzione.....	10
Azienda Ospitante.....	13
Struttura della Tesi.....	15
Capitolo 1.....	16
Ricerca Bibliografica.....	16
1.1 Formulazione dell'obiettivo della ricerca.....	16
1.2 Strumenti di consultazione e Keywords per la ricerca.....	17
1.3 Formulazione delle query di ricerca.....	18
1.4 Estrazione e pulizia dei dati.....	19
1.5 Analisi dei Dati e Statistica Descrittiva.....	19
1.5.1 Confronto tra Banche Dati.....	19
1.5.2 Confronto tra Keywords.....	21
1.6 Text Mining degli Abstract.....	23
1.6.1 Text Mining.....	24
1.6.2 Dati dell'Analisi.....	24
1.6.3 Processo di Text Mining.....	25
1.6.4 Applicazioni del Text Mining.....	27
1.7 Conclusioni della Ricerca Bibliografica.....	30
1.7.1 Stato dell'arte del Multivariate Statistical Process Control.....	31
Processo considerato e Dati del Caso di Studio.....	33
2.1 Storia dell'azienda Mixer Spa.....	33
2.2 Prodotti e Clienti della Mixer Spa.....	34
2.3 Processi.....	36
2.4 Processo di Miscelazione.....	38
2.5 Approccio di miglioramento del processo.....	39
2.6 Dati disponibili.....	40
Capitolo 3.....	44
MSPC: Multivariate Statistical Process Control.....	44

3.1 Controllo Statistico Multivariato di Processo.....	44
3.2 Carte di Controllo Univariate: Problematiche e Confronto con la PCA.....	45
3.3 PCA: Analisi delle Componenti Principali.....	51
3.3.1 Scelta del numero di componenti per il modello.....	54
3.4 Costruzione del modello e delle Carte di controllo multivariate.....	55
3.5 Trattamento preliminare dei dati.....	58
3.6 Carta di Controllo Riassuntiva.....	58
3.7 Multiway-PCA.....	60
3.8 Limiti di Controllo delle Carte.....	63
3.9 Validazione del Modello e Calcolo delle Sensitività.....	65
3.10 Analisi Three Way dei dati: Modello Tucker 3.....	66
3.10.1 Modello Three Way Tucker3.....	67
3.10.2 MSPC Chart per il modello Tucker3.....	70
Capitolo 4.....	73
Applicazione delle Carte di Controllo Multivariate per il monitoraggio di un processo di Miscelazione.....	73
4.1 Multivariate Statistical Process Control.....	74
4.2 Dati di Processo.....	75
4.3 Costruzione del Modello e dei Limiti di Controllo delle Carte.....	76
4.4 Validazione delle Carte di Controllo.....	78
4.4.1 Validazione NOC delle Carte di Controllo.....	78
4.4.2 Validazione AOC delle Carte di Controllo.....	80
4.5 Carta di Controllo Riassuntiva.....	82
4.6 Individuazione delle Anomalie di Processo e Identificazione delle determinanti di lotti anomali.....	85
4.6.1 Parametri di processo determinanti della Variabilità Straordinaria. .	87
Capitolo 5.....	94
Monitoraggio online e Aggiornamento del modello.....	94
5.1 Simulazione degli Step di Controllo Statistico Online.....	95
5.1.2 I dati dei Lotti in input per Monitoraggio Online.....	95
5.1.3 Step della Simulazione di Controllo Statistico Online tra un lotto e il	

successivo.....	96
5.2 Impatto organizzativo e Risorse aziendali necessarie per l'introduzione del Sistema di Monitoraggio Online per il processo di Miscelazione.....	99
5.2.1 Impatto Organizzativo per lo sviluppo e l' attuazione del Sistema di Monitoraggio.....	99
5.2.2 Processi Direzionali.....	100
5.2.3 Processi Operativi.....	101
5.2.4 Gestione delle Risorse.....	101
5.2.5 Misurazione, Analisi e Miglioramento del Sistema di Controllo.....	103
5.3 Monitoraggio offline: Interpretazione delle carte di controllo.....	104
5.3.1 Diagnosi delle variabili di processo determinanti di anomalie.....	105
5.3.2 Monitoraggio di Processo offline.....	106
5.4 Aggiornamento del Modello MPCA.....	109
5.4.1 Aggiornamento Modello:Inserimento di lotti nel Set Training.....	110
Codice R.....	115
Bibliografia.....	126
Ringraziamenti.....	129

Indice delle Illustrazioni

Illustrazione 1: Marchio Kode.....	13
Illustrazione 2: Grafico a barre degli Articoli in ogni Banca-Dati.....	20
Illustrazione 3: Grafico a barre degli Articoli in ogni Banca-Dati.....	20
Illustrazione 4: Grafico a barre delle Riviste per ogni Keywords.....	21
Illustrazione 5:Grafico a barre degli Articoli per ogni Keywords.....	22
Illustrazione 6: HeatMap Keywords/Banche Dati.....	23
Illustrazione 7: Dendrogramma degli Articoli estratti dalla Banca Dati Pubmed. 28	
Illustrazione 8: Dendrogramma degli Articoli estratti dalla Banca Dati Acs.....	29
Illustrazione 9: Pellets in materiale plastico (Sito della Mixer).....	35
Illustrazione 10:Immagini di Cavi di Plastica (Sito della Mixer).....	35
Illustrazione 11: Diagramma a blocchi dei Processi Produttivi dell'Azienda Mixer Spa per la realizzazione di pellets di materiale plastico.....	36
Illustrazione 12: Diagramma a blocchi del Processo di Miscelazione di Tipo Batch dell'Azienda Mixer Spa.....	39
Illustrazione 13: Profilo temporale medio della Variabile della Temperatura della Mescola nei 4 istanti di campionamento.....	42
Illustrazione 14: Profilo temporale medio della Variabile della Durata della Sottofase nei 4 istanti di campionamento.....	42
Illustrazione 15: Profilo temporale della Variabile della Potenza necessaria per la miscelazione nei 4 istanti di campionamento.....	43
Illustrazione 16: Profilo temporale medio della Variabile del Peso della Mescola nei 4 istanti di campionamento.....	43
Illustrazione 17: Carta di controllo univariata di Shewhart.....	46
Illustrazione 18: Carte di controllo univariata per un parametro di processo X1 con campione rosso rappresentativo del lotto non conforme all'interno dei limiti di confidenza rappresentati dalle linee rosse tratteggiate.....	48
Illustrazione 19: Carte di controllo univariata per un parametro di processo X2 con campione rosso rappresentativo del lotto non conforme all'interno dei limiti	

di confidenza rappresentati dalle linee rosse tratteggiate.....	48
Illustrazione 20: Rappresentazione Congiunta di due Variabili di Processo.....	50
Illustrazione 21: Piano definito dalle componenti principali PC1 e PC2.....	53
Illustrazione 22: Piano definito dalle componenti principali PC1 e PC2 circoscritto dai limiti di controllo e rappresentazione di lotti fuori norma per entrambe le carte di controllo (campioni in rosso)(Wise et al., 1997).....	57
Illustrazione 23: Carta Riassuntiva delle carte di controllo e Q.....	59
Illustrazione 24: Unfolding della matrice Three-Way nella direzione dei Batches I, in una matrice bidimensionale di dimensioni $I \times JK$ (Lotti x Variabili/Tempo) (Gunther et al., 2007).....	61
Illustrazione 25: Array di dati Tridimensionali e Decomposizione Multyway PCA (Wise et al., 1997).....	62
Illustrazione 26: Proiezione dei dati reali sui tre modelli di riferimento di Tucker3 modello A,modello B e modello C (Acar et al., 2007).....	68
Illustrazione 27: Dataset analizzati nel caso studio: dataset “Training”, dataset “Test NOC” e dataset “Test AOC”	75
Illustrazione 28: Risultati del Test di Validazione delle Carte: Carta di controllo di Hotelling del Set di lotti Test NOC (lotti in condizioni operative normali).....	79
Illustrazione 29: Risultati del Test di Validazione delle Carte: Carta di controllo Q del Set di lotti Test NOC (lotti in condizioni operative normali).....	79
Illustrazione 30: Risultati del Test di Validazione delle Carte: Carta di controllo di Hotelling del Set di lotti Test AOC (lotti in condizioni operative anormali).....	81
Illustrazione 31: Risultati del Test di Validazione delle Carte: Carta di controllo Q del Set di lotti Test AOC (lotti in condizioni operative anormali).....	81
Illustrazione 32: Risultati del Test di Validazione delle Carte: Carte di Controllo Riassuntiva del Set di lotti “Training”(cerchi neri), del Set di lotti Test NOC (lotti in condizioni operative normali,cerchi verdi), e del Set di lotti Test AOC (lotti in condizioni operative anormali,cerchi rossi).....	83
Illustrazione 33:Risultati della simulazione di monitoraggio su dati di processo forniti dall'azienda Mixer Spa: Carta di Controllo di 120 lotti.....	86
Illustrazione 34:Risultati simulazione monitoraggio sui dati forniti dall'azienda	

Mixer Spa: Carta di Controllo di Hotelling di 120 lotti.....	87
Illustrazione 35: Contribution Plot del Lotto 53-esimo classificato anomalo.....	89
Illustrazione 36: Contribution Plot del Lotto 56-esimo classificato anomalo.....	89
Illustrazione 37: Contribution Plot del Lotto 54-esimo classificato anomalo.....	90
Illustrazione 38: Carta di controllo di lotti monitorati in ordine successivo nel tempo con un andamento ciclico indicativo del fuori controllo del processo (fenomeno chiamato Deriva).....	91
Illustrazione 39: Proiezione delle variabili di processo originarie nello spazio dimensionale definito dalle prime due componenti principali del modello MPCA costruito dal dataset di lotti Training.....	92
Illustrazione 40: Step del Controllo Online di Processo.....	97
Illustrazione 41: Pianificazione del Sistema di monitoraggio per il Processo di Miscelazione.....	100
Illustrazione 42: Sviluppo del Sistema di Monitoraggio: Database e Interfacce	102
Illustrazione 43: Scenario 1: Carta di Controllo di un set di lotti in normali condizioni operative.....	107
Illustrazione 44: Scenario 2: Carta di Controllo di un set di lotti con trend positivo indicativo di anomalia.....	107
Illustrazione 45: Scenario 3: Carta di controllo di un set di lotti con la presenza di un solo un lotto anomalo.....	108
Illustrazione 46: Scenario 4: Carta di controllo rappresentativa di un'anomalia di Processo.....	108
Illustrazione 47: Step di monitoraggio offline e Aggiornamento del modello...	111
Illustrazione 48: Sottofasi del processo di Aggiornamento del modello.....	112

Indice delle tabelle

Tabella 1: Variabili del Processo di Miscelazione misurate dai sensori.....	41
Tabella 2: Sensitività: Risultati del Test NOC (Lotti in Condizioni Operative Normali) per verificare le performance delle Carte di Controllo costruite in base al Modello MPCA.....	80
Tabella 3: Sensitività AOC: Risultati del Test AOC (Lotti Condizioni Operative Anormali) per verificare le performance delle Carte di Controllo costruite in base al Modello MPCA.....	82
Tabella 4: Matrice dei dati riferita al nuovo lotto i-esimo, in input alle carte, per il monitoraggio online: dimensioni 1x16 (4 variabili x 4 istanti).....	96

Introduzione

In questa tesi si discute l'applicazione delle tecniche di Controllo Multivariato di Processo ai fini del miglioramento della qualità in ambito industriale.

Il controllo Multivariato di Processo è una tecnica statistica finalizzata al miglioramento della qualità, che insieme alle attività operative e gestionali assicura che il prodotto rispetti le specifiche pianificate. Un processo deve essere controllabile, ripetibile e stabile per realizzare un prodotto rispondente alle specifiche. L'applicazione dello Statistical Process Control provoca una riduzione dei costi di produzione e un'ottimizzazione della linea produttiva monitorando e migliorando il processo. Le carte di controllo Univariate sono state introdotte da Walter Shewhart. Il principio su cui si basano le carte di controllo Univariate è quello di segnalare un allarme nel momento in cui la grandezza misurata superi i limiti di controllo. Le suddette carte monitorano singolarmente ogni parametro di processo che si reputa incidere sulla qualità del prodotto, non considerando le correlazioni presenti tra le variabili. Nell'ambito industriale contemporaneo, però, il monitoraggio di diverse caratteristiche di processo avviene simultaneamente, perché i parametri di processo sono correlati tra loro e vanno ad incidere sulla qualità del prodotto, quindi il controllo aggregato di tali caratteristiche risulta essere indispensabile. Il monitoraggio di processo genera delle informazioni sul processo produttivo, attraverso l'acquisizione e l'analisi dei dati che vengono misurati da sensori posti lungo la linea produttiva, realizzando il monitoraggio e garantendone la stabilità di processo. In questa tesi è proposto l'approccio di Controllo Statistico

Multivariato per il Monitoraggio di Processo Online che tiene in considerazione, contemporaneamente, tutte i parametri che descrivono e caratterizzano un processo, e quindi anche le informazioni di correlazione tra le caratteristiche di qualità. I tipici dati provenienti dai processi batch richiedono lo studio delle traiettorie nel tempo delle variabili misurate per tutta la durata di produzione di un lotto (Nomikos & MacGregor, 1995). Se sono disponibili nei database aziendali i dati storici su molti lotti, possono essere sviluppati dei modelli Multivariati basati sul metodo dell'Analisi delle Componenti Principali (Jackson, Mudholkar, 1979). Tale approccio fornisce informazioni inerenti allo stato del processo (in or out-control), una diagnosi delle determinanti di anomalie riscontrate sui lotti e informazioni relative alle correlazioni tra variabili di processo. L'Analisi delle Componenti Principali (PCA) effettua una riduzione delle dimensioni della matrice dei dati proiettando le variabili di processo in uno spazio ridotto, costruito su variabili latenti in grado di mantenere tutto il contenuto informativo utile del sistema.

I metodi di controllo multivariato proposti si basano sulle carte T^2 di Hotelling, e le carte di controllo Q, costruite utilizzando la tecnica della Multiway Principal Component Analysis (MPCA), che è un'estensione della classica PCA, la quale permette una riduzione dimensionale dei dati di processo. E' proposto inoltre un approccio alternativo alla MPCA, che utilizza il modello Three-Way di Tucker, detto anche Tucker3 (Van Sprang, Ramaker, Westerhuis, Smilde, & Wienke, 2005). L'obiettivo della tesi è quello di fornire un'applicazione della tecnica di Controllo Statistico Multivariato di Processo, con lo scopo di costruire un modello matematico in grado di descrivere le relazioni tra le variabili di processo e le proprietà di un lotto in riferimento ai requisiti specificati. La metodologia sviluppata è stata pensata per intervenire online tra un lotto e il successivo, classificando un lotto come conforme o meno. Il caso di studio che è stato analizzato, riguarda i dati di processo forniti dall'Azienda Mixer Spa, produttrice

di Pellets in materiale plastico per la realizzazione di cavi in plastica. I dati sono relativi ad un Processo di Miscelazione di tipo batch.

Nel Capitolo 1 è presentata la Ricerca Bibliografica che è stata svolta, per conoscere lo Stato dell'Arte del "Multivariate Statistical Process Control" attraverso l'analisi della letteratura scientifica che ne evidenzia i campi applicativi. Nel Capitolo 2 è presente una descrizione del processo produttivo e dei dati di processo, che sono stato utilizzati per la costruzione del sistema di monitoraggio multivariato di processo. Nel Capitolo 3 sono descritti due approcci per la costruzione delle Carte di Controllo, partendo da una matrice di dati Three-Way: il modello della Multiway Analisi delle Componenti Principali e il modello Tucker 3, le carte di controllo T^2 di Hotelling e le carte di controllo Q. Nel Capitolo 4 è presente l'applicazione ai dati di processo dell'azienda Mixer Spa. Il modello di monitoraggio è stato costruito per la prima fase produttiva di tipo batch, ovvero il processo di Miscelazione. Nel Capitolo 5 è proposto un metodo di implementazione del modello sia per il monitoraggio online tra un batch e il successivo e sia per un monitoraggio post-produzione. Infine sono presentate le conclusioni, la bibliografia e il codice che è stato implementato tramite il linguaggio di programmazione statistica R per la costruzione del modello e delle carte di controllo.

Azienda Ospitante

Il lavoro di tesi nasce da un tirocinio della durata di sei mesi svolto presso l'azienda Kode Srl.



Illustrazione 1: Marchio Kode

Kode è una giovane Società di Consulenza Scientifica nel campo del Data Science. E' costituita da un gruppo di esperti in diversi settori: Statistica Multivariata, Data Mining, Chemiometria, Chemoinformatica e Ricerca Socio-Economica. Fornisce supporto ad aziende, istituzioni ed enti di ricerca: attraverso l'analisi dei loro dati, valorizzando il loro know-how con strumenti analitici, sviluppo di modelli, visualizzazioni e produzioni di report e di work-flow. Le attività svolte dall'azienda sono le seguenti:

- **Analisi e Elaborazione di dati**

Data Mining per ogni tipo di dati e applicazioni di tecniche chemiometriche per ottimizzare i processi industriali.

- **Chemoinformatica**

Gli strumenti più avanzati per gestire l'informazione chimica. Organizzazione dei dataset, elaborazione delle molecole, modelli QSAR e QSPR per la predizione delle proprietà.

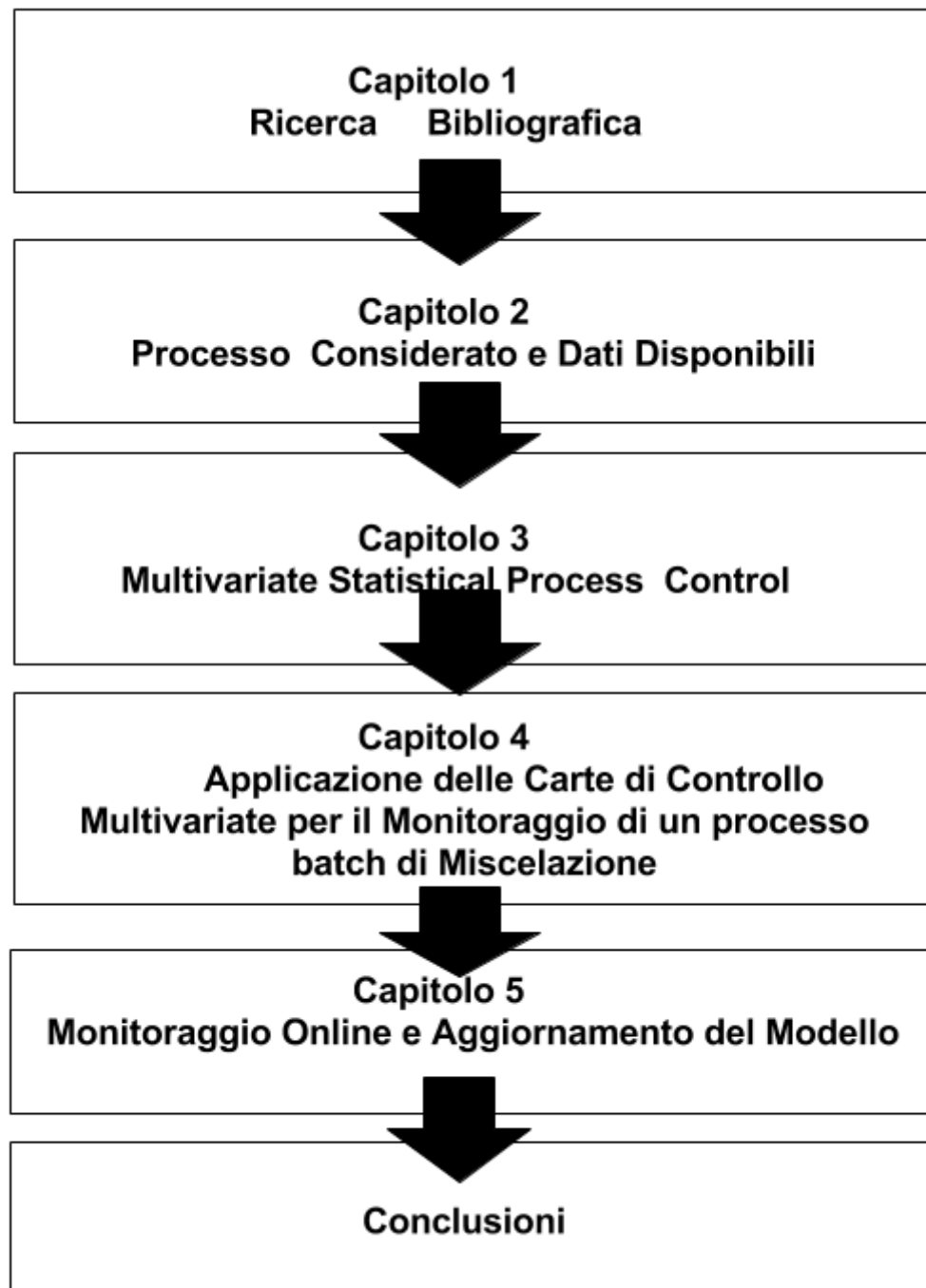
- **Open Data**

Kode propone strategie innovative per sfruttare le potenzialità degli Open Data.

- **Formazione**

Offre formazione su i temi più avanzati dell'analisi e dell'elaborazione dei dati dai corsi sul linguaggio di programmazione R alle strategie open data.

Struttura della Tesi



Capitolo 1

Ricerca Bibliografica

Il Controllo di Processo Multivariato è uno degli argomenti più in rapido sviluppo delle aree di controllo statistico, pertanto è stato indispensabile effettuare una Ricerca Bibliografica per conoscere lo Stato dell'Arte della Letteratura Scientifica di tale argomento e per comprendere la trasversalità dell'ambito di applicazione. La Ricerca bibliografica è articolata nelle seguenti fasi: formulazione dell'obiettivo della ricerca, identificazione degli strumenti di consultazione, delle keywords per le query di ricerca, scelta dell'intervallo temporale di analisi, formulazione delle query di ricerca, estrazione e pulizia dei dati, analisi dei dati, statistica descrittiva, text mining, conclusioni e sintesi dello Stato dell'Arte.

1.1 Formulazione dell'obiettivo della ricerca

L'obiettivo di tale ricerca è la conoscenza dello Stato dell'Arte della letteratura scientifica inerente al Controllo di Processo Multivariato e dei diversi ambiti d'applicazione. Il lavoro di text mining svolto all'interno della ricerca bibliografica ha lo scopo di classificare gli articoli scientifici estratti dalle Banche Dati online, al fine di ottenere dei cluster omogenei per comprendere le diverse tecniche utilizzate nel Controllo Statistico di Processo e i campi in cui queste tecniche trovano applicazione.

1.2 Strumenti di consultazione e Keywords per la ricerca

Gli strumenti di consultazione utilizzati sono Banche Dati bibliografiche online nello specifico:

- Emerald
- Sciencedirect
- Acs Publications
- Pubmed
- Wiley

Emerald Group Publishing Limited è una casa editrice di Riviste Accademiche in materia di gestione, commercio, formazione e ingegneria.

ScienceDirect è un sito web gestito dalla casa editrice anglo-olandese Elsevier. Si tratta di una piattaforma per l'accesso a circa 2.500 riviste accademiche e oltre 26.000 e-book. Le riviste sono raggruppate in quattro sezioni principali: Scienze fisiche e Ingegneria, Scienze della vita, Scienze della salute, Scienze sociali e umane.

ACS publications è un sito online gestito dalla società scientifica American Chemical Society con sede negli Stati Uniti che sostiene la ricerca scientifica nel campo della Chimica. E' la più grande società scientifica del mondo e una delle fonti principali di informazioni scientifiche autorevoli. La sua divisione di pubblicazioni produce decine di riviste accademiche, tra cui il prestigioso Journal of American Chemical Society.

Pubmed è un database bibliografico contenente informazioni sulla letteratura Scientifica Biomedica dal 1949 ad oggi.

Wiley Online Library è un sito online gestito dalla casa editrice Wiley

specializzata a livello mondiale in pubblicazioni accademiche che commercializza i suoi prodotti a professionisti e consumatori, studenti, docenti nel campo dell'istruzione superiore, ricercatori e professionisti in discipline scientifiche, tecniche, mediche. La tipologia dei documenti scelti sono articoli di riviste scientifiche. L'arco temporale considerato copre un periodo di 10 anni che va dal 2005 al 2015.

L'ambito di ricerca è circoscritto da 8 keywords :

- MSPC,
- Multivariate Statistical Process Control;
- Statistical Process Control;
- Process Control;
- Process Monitoring;
- Quality Control;
- Quality Monitoring;
- Raw Materials.

1.3 Formulazione delle query di ricerca

Il lavoro bibliografico è condotto ricercando ogni keywords all'interno dell'abstract, delle parole chiave o del titolo di ogni articolo scientifico, congiuntamente al criterio di selezione "PCA" oppure "PLS" all'interno dell'abstract. L'uso dei Metodi di Proiezione della PCA (Analisi delle Componenti Principali) e della PLS (Regressione dei Minimi Quadrati Parziali) è fondamentale nell'ambito della Statistica Multivariata.

1.4 Estrazione e pulizia dei dati

I dati degli articoli scientifici provenienti dalle banche dati Online, sono stati elaborati utilizzando il programma OpenRefine che ha reso possibile la loro pulizia. Per ogni articolo scientifico sono state estratte le seguenti informazioni: titolo dell'articolo, titolo della rivista, autori, anno di pubblicazione, url, doi, issn, affiliation, abstract, referenze e keywords.

1.5 Analisi dei Dati e Statistica Descrittiva

I dati provenienti dalle query sono caricati e analizzati tramite R, ovvero un linguaggio di programmazione per il calcolo statistico. Sono state create due tipologie di matrici di dati:

- la prima matrice è composta da quattro colonne identificative delle seguenti variabili: keywords, banca dati, rivista scientifica e conteggio delle riviste associata alla keywords.
- la seconda matrice è composta da quattro colonne identificative delle seguenti variabili: keywords, banca dati, rivista scientifica e conteggio degli articoli.

Le matrici sono state elaborate per effettuare i confronti di seguito riportati.

1.5.1 Confronto tra Banche Dati

Le illustrazioni 1 e 2 indicano per ciascuna banca dati il numero di riviste e di articoli trovati per tutte le keywords. La banca dati Sciencedirect presenta il

numero di occorrenze più elevato, intorno alle 300 riviste scientifiche e articoli.

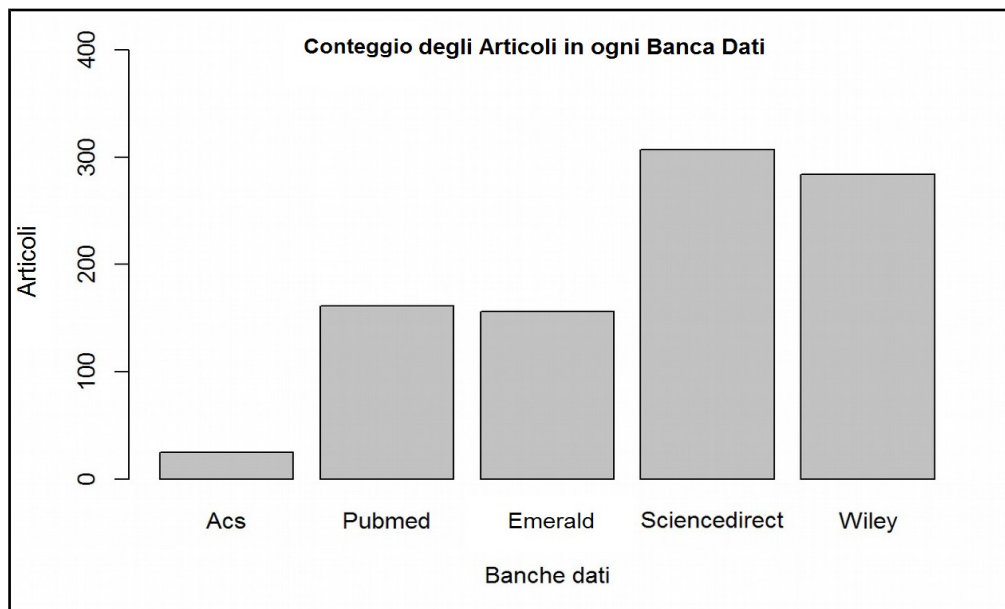


Illustrazione 2: Grafico a barre degli Articoli in ogni Banca-Dati

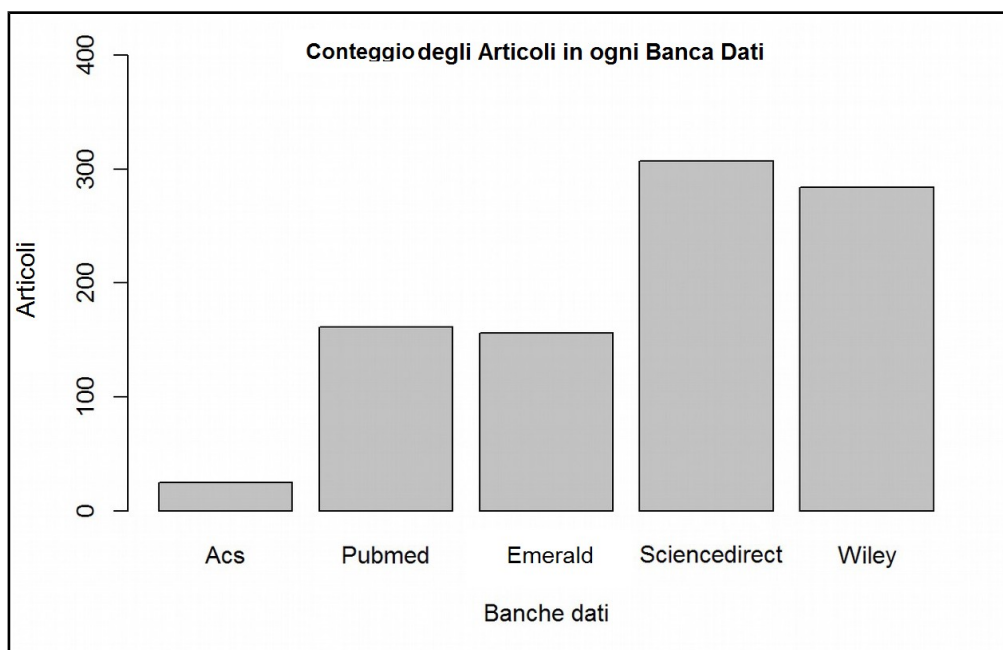


Illustrazione 3: Grafico a barre degli Articoli in ogni Banca-Dati

1.5.2 Confronto tra Keywords

Le illustrazioni 5 e 6 mostrano per ogni keywords, il numero di riviste e di articoli presenti nelle banche dati. Le keywords “Statistical Process Control” e “Process Monitoring” presentano il numero di occorrenze più alto compreso tra le 150 e le 200 Riviste Scientifiche , e tra i 1000 e 1200 Articoli Scientifici.

Legenda:

- a) MSPC
- b) Multivariate Statistical Process Control
- c) Statistical Process Control
- d) Process Control
- e) Process Monitoring
- f) Quality Control
- g) Quality Monitoring
- h) Raw Materials

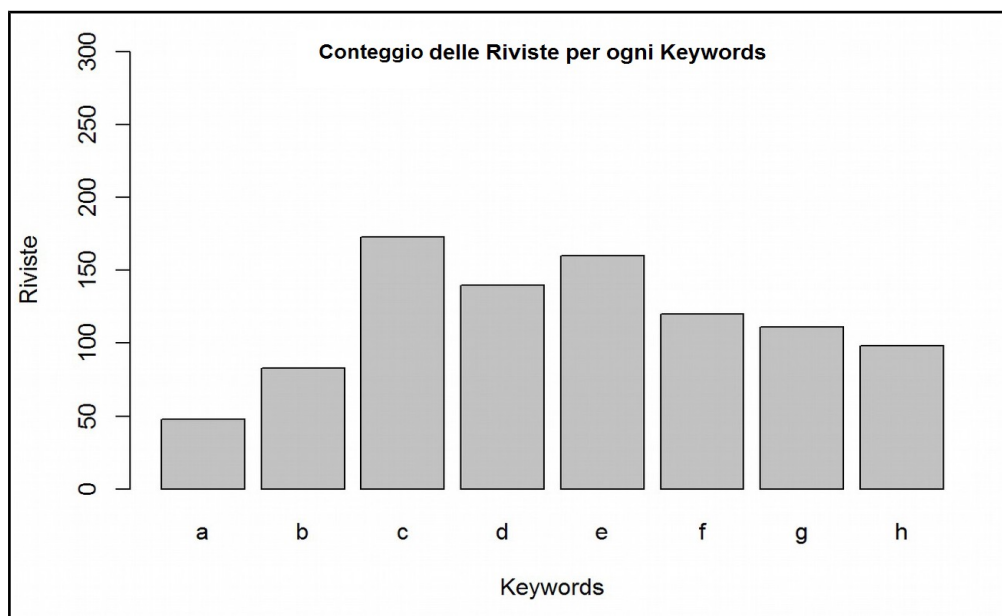


Illustrazione 4: Grafico a barre delle Riviste per ogni Keywords

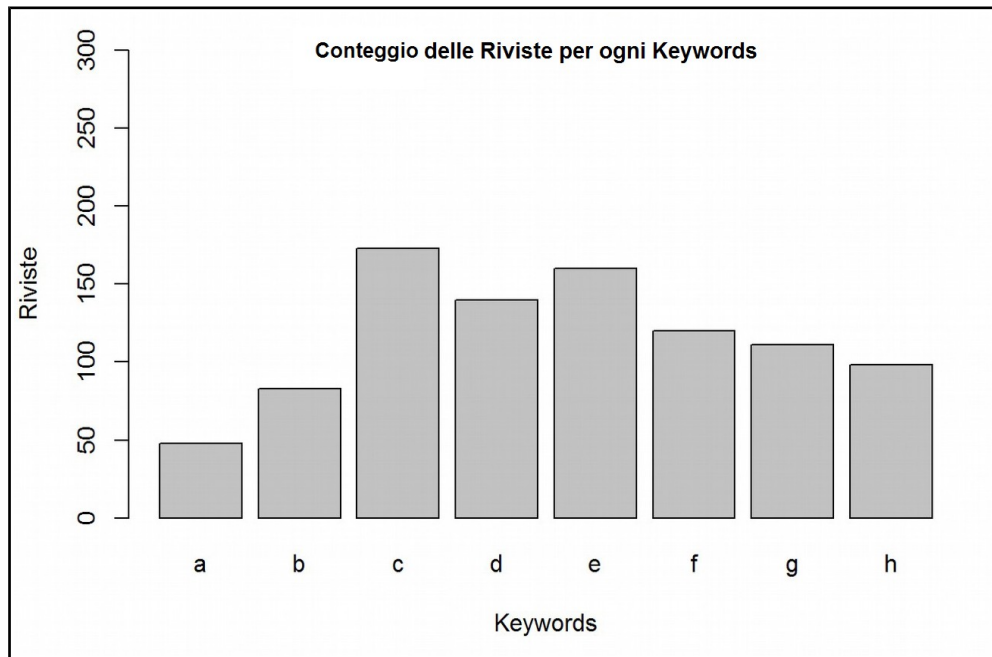


Illustrazione 5: Grafico a barre degli Articoli per ogni Keywords

1.5.3 Confronto tra Keywords per singola banca dati

L'illustrazione seguente è un heatmap, che mette in correlazione le banche dati agli articoli trovati per ogni keywords, le zone rosso identificano le ricerche più ricche di articoli all'interno di una banca dati e le zone bianche le meno ricche.

Legenda:

- a) MSPC
- b) Multivariate Statistical Process Control
- c) Statistical Process Control
- d) Process Control
- e) Process Monitoring

- f) Quality Control
- g) Quality Monitoring
- h) Raw Materials

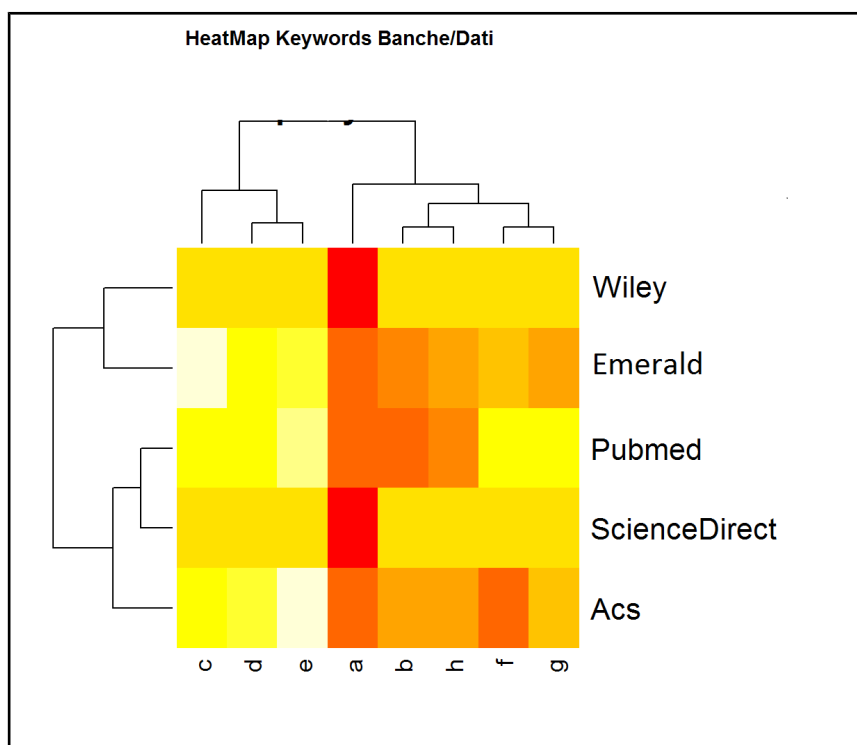


Illustrazione 6: HeatMap Keywords/Banche Dati

1.6 Text Mining degli Abstract

L'obiettivo del lavoro Text Mining svolto è classificare i documenti simili così da poter individuare gli articoli più significativi per l'argomento di tesi "Multivariate Statistical Process Control" e conoscere i diversi ambiti applicativi di questa tecnica presenti allo Stato dell'Arte.

1.6.1 Text Mining

Con il termine "Text Mining" si parla di un esempio particolare di Data Mining, un'elaborazione automatizzata di un insieme di documenti testuali, volta ad effettuare una classificazione degli elementi che ne fanno parte in base all'argomento trattato o ad altre caratteristiche distintive. Il programma che è stato utilizzato per l'elaborazione dei dati è Rstudio. Rstudio è uno dei principali e più versatili ambienti di sviluppo Open-Source per l'analisi statistica dei dati. L'ambiente R è basato sul concetto di "package". Un package è un insieme di strumenti che svolgono determinate funzioni, ma può contenere anche solo dati o solo documentazione. Attualmente è disponibile un'ampia gamma di packages online utilizzabili per la risoluzione di specifici problemi.

Il package Tm permette la gestione di documenti di testo, la manipolazione dei documenti e facilita l'utilizzo di formati di testo eterogenei in R. Il pacchetto supporta l'esportazione di collezioni di matrici termine-documento che possono essere facilmente costruiti da documenti di testo. Tutte queste funzionalità consentono all'utente di eseguire indagini di text mining utilizzando i metodi statistici forniti da Rstudio.

1.6.2 Dati dell'Analisi

I dati analizzati comprendono gli Abstract degli articoli scientifici pubblicati tra 2005 e 2015 in lingua inglese estratti dalle banche dati online Pubmed, ScienceDirect, Emerald e Acs, utilizzando come Keywords per le query di ricerca online i seguenti termini: Multivariate Statistical Process Control, Statistical Process Control, Process Control, Process Monitoring, Quality Control, Quality Monitoring e Raw Materials.

1.6.3 Processo di Text Mining

Il processo di text mining consta delle seguenti fasi:

- Preprocessing
- Hierarchical Clustering
- Caratterizzazione dei Clusters

- **Preprocessing**

Gli Abstract degli articoli scientifici, precedentemente estratti dalle banche dati online sono importati sul programma Rstudio e sono stati analizzati con la libreria per il text mining, tm. Una delle prime operazioni da compiere è la raccolta di tutti i documenti in un unico Corpus di testo.

La fase di Stemming consiste nel processo di cancellazione del suffisso dei termini, per recuperare le loro radici. Si tratta di una tecnica comunemente utilizzata nella ricerca di testi, in quanto riduce la complessità del testo senza alcuna perdita grave delle informazioni per le applicazioni tipiche.

I package utilizzati in R sono Rstem e Snowball, la funzione principale è wordStem () che utilizza l'algoritmo di Porter di stemming.

Altri due fasi comuni di pre-elaborazione, sono la rimozione dello spazio bianco nel Corpus, la conversione in minuscolo e la rimozione di termini detti Stopwords. Le Stopwords sono parole che si ripresentano spesso nel linguaggio comune, ma hanno valore quasi nullo ai fini del lavoro di Text Mining, perciò sono eliminate. Successivamente viene creata una matrice termine-documento depurata delle Stop-words e una lista contenente l'elenco delle Stopwords eliminate. Al termine della pre-elaborazione, la matrice avente come righe i termini significativi estratti dal corpus, e come colonne gli ID degli Articoli è ulteriormente ridotta eliminando i termini considerati di valore nullo per

l'analisi, lasciando quelli con frequenza più elevata e con valore significativo come ad esempio gli ambiti di applicazione delle tecniche sotto esame (Multivariate Statistical Process Control, Statistical Process Control, Process Control, Process Monitoring, Quality Control, Quality Monitoring, Raw Materials).

- **Hierarchical Clustering**

L'obiettivo della Cluster Analysis è quello di identificare gli articoli della matrice termine-documenti collegati ad un argomento comune ed inserirli all'interno di clusters omogenei. Le variabili della matrice termini-documenti sono di tipo qualitativo, perciò la matrice è stata trasformata in binaria. Per una variabile qualitativa (ovvero i termini) che vale 1 nel caso della sua presenza e 0 nel caso di assenza, è definibile una distanza, pertanto la variabile qualitativa viene trasformata in quantitativa. Per proseguire l'analisi, occorre stabilire una misura di diversità per calcolare la matrice delle distanze. Prima del calcolo delle distanze, la matrice termini-documenti è stata standardizzata. Si è scelto di usare la distanza della correlazione di Pearson, che varia tra -1 e 1 e misura la relazione lineare tra 2 vettori. Il metodo Ward è stato utilizzato come algoritmo gerarchico aggregativo. Tale metodo è diretto alla minimizzazione della varianza all'interno dei gruppi. Ad ogni passo questo algoritmo tende ad ottimizzare la partizione ottenuta tramite l'aggregazione di due elementi. Una partizione si considera tanto migliore, quanto più le classi risultano omogenee al loro interno e differenti l'una dall'altra. Nel clustering gerarchico vengono creati più livelli, generati da applicazioni ricorsive dell'algoritmo. Il set iniziale di documenti viene diviso in più Clusters, ognuno dei quali viene nuovamente raggruppato ricorsivamente, fino al soddisfacimento di una condizione prefissata di omogeneità perciò non è necessario definire a priori la struttura dell'albero.

- **Caratterizzazione dei clusters**

Il metodo di analisi per dare una caratterizzazione al cluster si basa sulla valutazione count-based. In questo metodo si contano le frequenze dei termini più alte in ogni cluster, che sono significative per associare al gruppo di articoli un argomento comune.

1.6.4 Applicazioni del Text Mining

L'analisi di Text mining degli Abstract, è stata applicata alle seguenti banche dati Online: Pubmed, ScienceDirect, Acs, Emerald.

La procedura di Text Mining seguita è la seguente:

1. creazione del corpus di dati e fase di preprocessing;
2. creazione della matrice di termini-documenti;
3. calcolo della matrice delle distanze e Clustering tramite il metodo Ward;
4. caratterizzazione dei cluster tramite conteggio delle occorrenze.

- **Text Mining in Pubmed**

Gli articoli scientifici estratti dalla banca dati online Pubmed, sono 213. Questi articoli sono stati suddivisi in 5 clusters. L'argomento principale del primo cluster è la tecnica della PCA utilizzata nel campo chemiometrico, biomedico, metabolomico, geostatistico, e del software. Nel secondo cluster gli articoli scientifici si concentrano sul Controllo di Processo, utilizzando sia l'Analisi delle componenti principali e la Partial Least Square Regression, nell'ambito biochimico, socioeconomico, della produzione tessile e dei network. Nel terzo cluster gli articoli trattano delle tecniche sia di PCA e di PLS, per il controllo chemiometrico di processo degli impianti batch. Nel quarto cluster gli articoli trattano la tecnica di Partial Least Square, per il controllo di processo di impianti

batch, nel beverage manufacturing e in ambito economico. Il quinto cluster si concentra sull'argomento di tesi “Multivariate Statistical Process Control”, applicato al monitoraggio di processo ,al campo microbiologico, dell'healthcare, del beverage manufacturing, della metabolomica. Nell'illustrazione seguente, è riportato il dendrogramma degli articoli estratti dalla banca dati Pubmed, dove è visibile la suddivisione dei suddetti in 5 clusters, evidenziati in rosso.

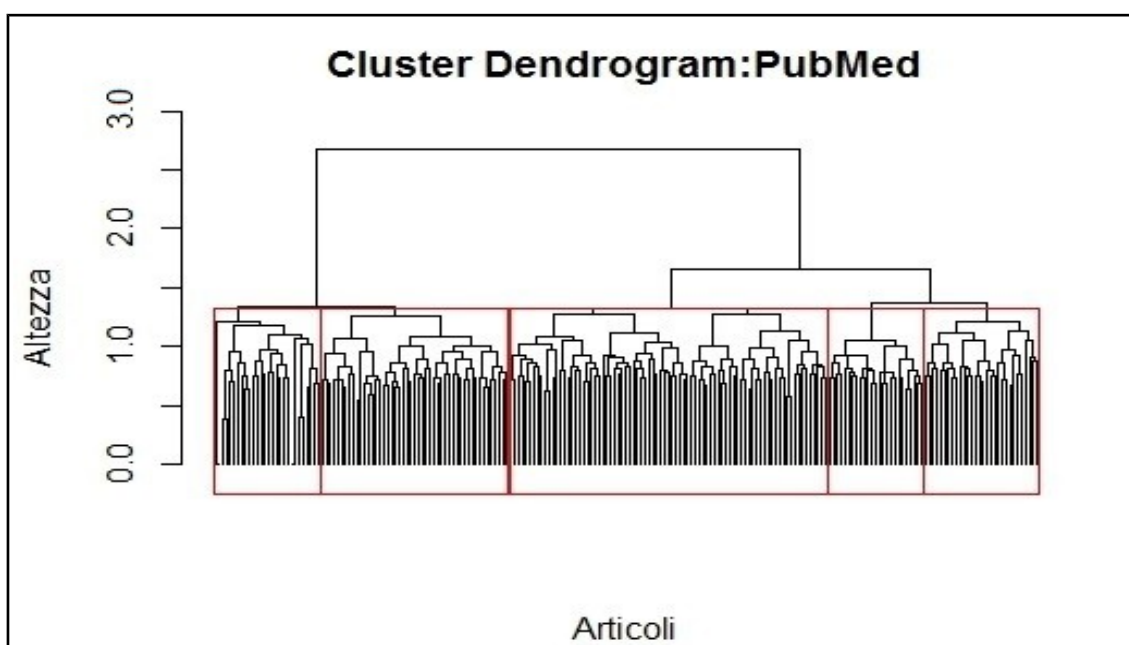


Illustrazione 7: Dendrogramma degli Articoli estratti dalla Banca Dati Pubmed

• Text Mining in ScienceDirect

Gli articoli scientifici estratti dalla banca dati Online Science Direct sono 449 e sono stati suddivisi in 5 clusters. Nel primo cluster i termini più ricorrenti sono fault, monitoring e PCA; nel secondo cluster PCA, PLS e regressione; nel terzo cluster monitoring, batch, PLS; nel quarto cluster fault, monitoring, least, PCA,

pharmaceutical; nel quinto cluster i termini più ricorrenti sono: industry, monitoring, chemometrics, business, PCA, PLS, manufacturing, validation, technology; nel sesto cluster: sensory, industry, monitoring, PCA e PLS.

• Text Mining in Acs

Gli articoli scientifici estratti dalla banca-dati online Science Direct sono 77 articoli e sono stati suddivisi in 4 clusters. Il primo cluster presenta i seguenti termini con occorrenza massima: PCA, quality, production, batch; Il secondo cluster: PCA, PLS, MSPC, batch, food; Il terzo cluster: electrochemical, pls, technology, sensitivity. Il quarto cluster: chemometrics, batch, operations, industry. Nell'illustrazione seguente, è riportato il dendrogramma degli articoli estratti dalla banca dati Acs, dove è visibile la suddivisione in 4 clusters, evidenziati in rosso.

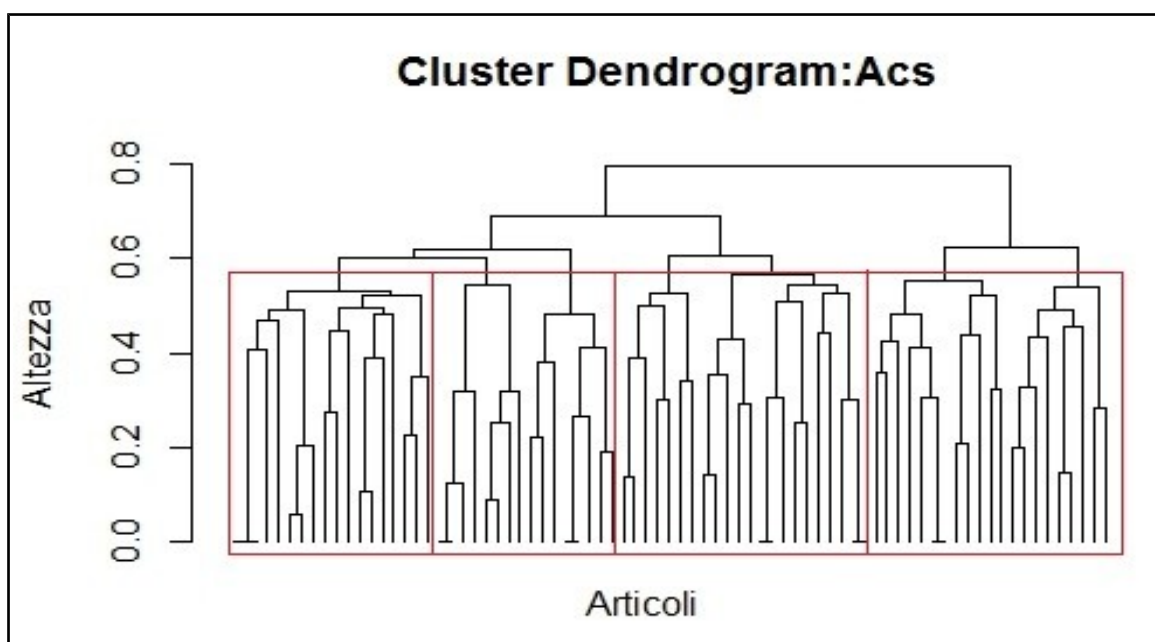


Illustrazione 8: Dendrogramma degli Articoli estratti dalla Banca Dati Acs

- **Text Mining in Emerald**

Gli articoli scientifici estratti dalla banca dati online Emerald sono 120 e sono stati divisi in 5 clusters. Nel primo cluster i termini di maggiore occorrenza sono: control, PCA, fault, control; nel secondo cluster: control, mechanical, PCA, process; nel terzo cluster: control, machine, learning, PLS, PCA; nel quarto cluster i termini di maggiore occorrenza sono: control, PCA, monitoring, quality, process, regression, MSPC, fault; nel quinto cluster sono: PCA, monitoring, engineering e material.

1.7 Conclusioni della Ricerca Bibliografica

La suddivisione in clusters degli articoli per ciascuna banca-dati e la successiva caratterizzazione di ogni cluster, tramite il conteggio delle occorrenze dei termini, hanno evidenziato i clusters significativi per l'argomento di tesi "Multivariate Statistical Process Control", ovvero i gruppi di articoli che applicano tale approccio nell'ambito industriale per il controllo di processo batch. A fronte dei risultati di questa ricerca è stato deciso di approfondire lo studio degli articoli presenti nel quarto e quinto clusters del database ScienceDirect; tali articoli sono stati scelti perché connotati dai seguenti termini di maggior occorrenza: industry, monitoring, chemometrics, business, PCA, PLS, manufacturing, validation, technology e sensory.

La ricerca focalizzata all'interno dei gruppi creati e selezionati, ha permesso di individuare gli articoli scientifici sia teorici e applicativi su cui è stata costruita la tesi.

1.7.1 Stato dell'arte del Multivariate Statistical Process Control

Analizzando la letteratura che tratta la tematica del Controllo Statistico Multivariato, si evidenzia come la sua applicazione sia trasversale ed interessi i campi della Medicina, il mondo dei Software, la Chimica, l'Ambito Industriale, l'ambito della Sensoristica etc. Le tecniche statistiche alla base dell'approccio sono quelle dell'Analisi delle Componenti Principali (PCA) e della Regressione Multivariata (PLS). Tali tecniche si propongono di studiare l'andamento di serie di dati, definendo un modello statistico che cerchi di descrivere il comportamento di un fenomeno che debba essere tenuto sotto controllo. L'utilizzo di metodi a variabili latenti ha rivoluzionato il controllo di processo. Le carte di controllo possono essere considerate come indici di prestazione del processo. Permettono di rilevare precocemente guasti e fornire anche degli strumenti di diagnosi del problema. L'applicazione dei metodi di analisi multivariata sono stati estesi anche per il monitoraggio online (Kourti, 2005). La metodologia delle variabili latenti utilizza il database di processo, costituito da misurazioni di un numero di variabili molto elevato, ma essendo molto correlate tra loro, lo spazio in cui possono essere proiettate è ridotto. I metodi di Statistica Multivariata sono utili anche nel caso di altri problemi che riguardano i database industriali, ovvero il problema di dati mancanti. Approcci più recenti utilizzano i metodi della Multiway PCA e della Multiway PLS, utilizzati per gestire delle matrici tridimensionali (Kassidas, Macgregor, & Taylor, 1998). Un ipotesi per l'applicazione di queste tecniche che è stata introdotta da Nomikos e MacGregor, è che i lotti debbano avere la stessa durata. Nella realtà non è sempre così, soprattutto nei processi batch non completamente automatizzati, dove alcune fasi vengono lasciate a discrezione dell'operatore. Per gestire il problema della diversa durata dei lotti, sono presenti allo Stato dell'Arte diverse

tecniche che si occupano di Sincronizzazione, come la Dynamic Time Warping (DTW). Quest'ultima tecnica è utilizzata anche per il riconoscimento dell'insorgenza di una nuova fase durante lo svolgimento di un processo.

Capitolo 2

Processo considerato e Dati del Caso di Studio

In questo capitolo è presentato il caso studio che è stato affrontato nella tesi: il processo di miscelazione di un impianto di produzione di pellets in materiale plastico. I dati analizzati sono stati forniti dalla azienda Mixer Spa.

La Mixer è un'azienda produttrice di mescole di gomma e composti termoplastici per l'isolamento e il riempimento di cavi flessibili, cavi in gomma, cavi elettrici e cavi ignifughi. I prodotti della Mixer sono basati su polimeri e additivi senza alogeni e a bassa tossicità ed emissione a fumo leggero.

2.1 Storia dell'azienda Mixer Spa

Mixer ha iniziato la produzione dei cavi di plastica nel 1996, al fine di soddisfare le esigenze specifiche del settore fili e cavi. Grazie ad una forte crescita, l'azienda si è posizionata come un importante fornitore di mescole e composti ritardanti in fiamma, con una base di clienti che si estende in tutto il mondo. Mixer è il più grande fornitore di cavo flessibile, cavo di gomma, cavo di alimentazione e cavi

composti ritardanti in fiamma nel Mercato Italiano e possiede anche una parte importante del Mercato Spagnolo. Le linee di produzione completamente automatizzate permettono di produrre elevati volumi di prodotti. L'approccio dell'azienda Mixer è orientato verso il miglioramento continuo dei prodotti. Costantemente è presente una ricerca delle materie prime disponibili sul mercato, di soluzioni innovative che rispettino gli standard più rigorosi, per l'uso in applicazioni di nuovi prodotti. Da nuove miscele polimeriche e nuovi additivi, Mixer sviluppa pellets in materiale plastico all'avanguardia che hanno ciò che serve per soddisfare le esigenze in continua evoluzione del settore fili e cavi.

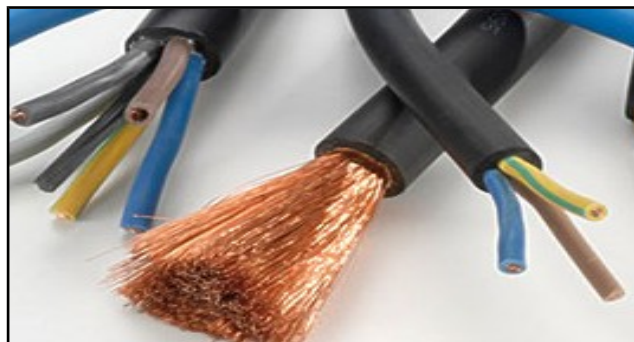
2.2 Prodotti e Clienti della Mixer Spa

Annualmente, oltre 20.000 tonnellate di composti per la produzione di cavi di plastica, fuoriescono dalle linee produttive dell'azienda. Mescolatori interni e estrusori vengono utilizzati per la produzione di pellets, ovvero cilindretti di composti termoplastici.

Nella compressione dei composti la temperatura aumenta, il composto è costretto ad attraversare dei fori conici di una trafilatura, che comprimono il materiale dando ad esso compattezza, mentre altre sostanze nel composto con la temperatura e pressione, si trasformano e svolgono la funzione di collante. In uscita dal foro della trafilatura che presenta la forma di un lungo filo, viene tagliato della misura desiderata ed in questo modo si ottengono dei cilindretti. Nella filiera produttiva la Mixer Spa si colloca come un'azienda fornitrice di pellets di materiale plastico per aziende produttrici di cavi in plastica.



*Illustrazione 9: Pellets in materiale
plastico (Sito della Mixer)*



*Illustrazione 10: Immagini di Cavi di
Plastica (Sito della Mixer)*

L'azienda Mixer Spa presenta la sua sede legale in Villa Prati di Bagnacavallo, in provincia di Ravenna. E' una società per azione che possiede un capitale sociale di 1.321.000 euro. I principali Clienti dell'azienda sono: Ariston Cavi, ElettroBrescia, Furas, General Cavi, Italian Cable Company. L'azienda Mixer offre anche il servizio di testing dei prodotti realizzati dai clienti, al fine di verificare la corretta reticolazione dei materiali al fine di determinare se i prodotti siano conformi alle norme in vigore.

2.3 Processi

La linea produttiva consta di due fasi:

- 1) Prima fase di tipo Batch
- 2) Seconda fase di tipo Continuo

Nell' Illustrazione 11, sono riportate i processi che compongono l'intera linea produttiva.

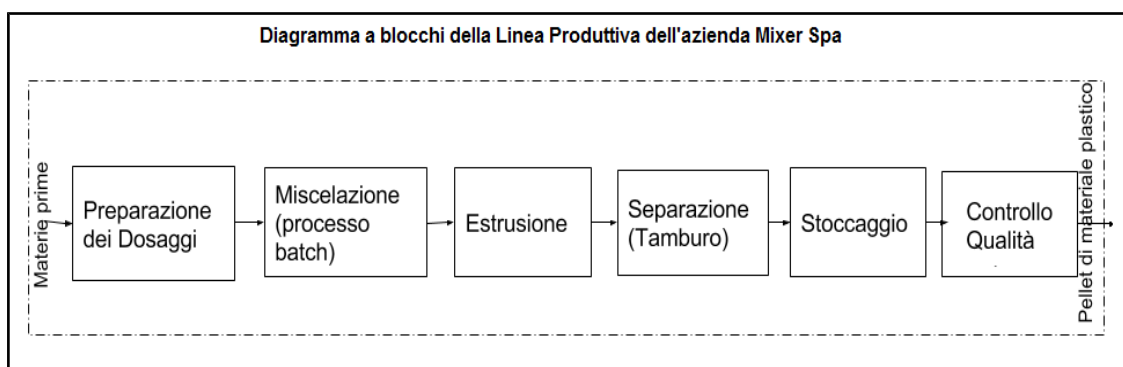


Illustrazione 11: Diagramma a blocchi dei Processi Produttivi dell'Azienda Mixer Spa per la realizzazione di pellets di materiale plastico

Di seguito è riportata una descrizione dei processi che compongono la linea produttiva di Mixer.

- 1. Preparazione dei dosaggi:** a monte della linea produttiva sono preparati i mix di dosaggi, costituiti da polimeri che andranno in input al processo di miscelazione. Le materie prime approvvigionate dalla Mixer sono sottoposte a controlli di qualità in ingresso.
- 2. Miscelazione:** un macchinario di miscelazione riceve in input un mix di

polimeri e di additivi, tramite lo svolgimento discontinuo del processo di miscelazione, che consta di quattro sottofasi, si giunge all'output di processo, ovvero una miscela polimerica.

3. **Estrusione:** la miscela polimerica è inserita all'interno di una tramoggia e compressa all'interno di una filiera. In uscita dal foro della filiera si ottiene il materiale estruso.
4. **Separazione:** un filtro a tamburo è utilizzato per la separazione di pellets dalla miscela polimerica precedentemente estrusa.
5. **Stoccaggio:** i pellets sono immagazzinati all'interno di big bag, contenitori flessibili di misure standard, utilizzati per immagazzinare grandi quantità di prodotti.
6. **Controllo di Qualità:** il laboratorio di qualità della Mixer, presente all'interno dello stabilimento produttivo, testa le proprietà dei Pellets, che sono stati stoccati all'interno dei big bag. La strumentazione e i test effettuati all'interno del laboratorio includono:
 - Analisi Termogravimetrica (TGA): tecnica che misura il peso di una sostanza, soggetta al riscaldamento con crescita controllata della temperatura.
 - Spettrometro a infrarossi: strumento utilizzato per la tecnica spettroscopica di assorbimento utilizzata per la caratterizzazione dei materiali.
 - Cromatografia liquida ad alte prestazioni (HPLC): tecnica cromatografica utilizzata per separare due o più composti presenti in un solvente, sfruttando l'equilibrio di affinità tra la "fase stazionaria" posta all'interno della colonna cromatografica e una "fase mobile" che fluisce all'interno di essa.
 - Prova di Trazione: prova di caratterizzazione del materiale, che consiste nel sottoporre un provino di dimensioni standard di un materiale in

esame, ad un carico inizialmente nullo che viene incrementato fino ad un valore massimo, che determina la rottura del materiale.

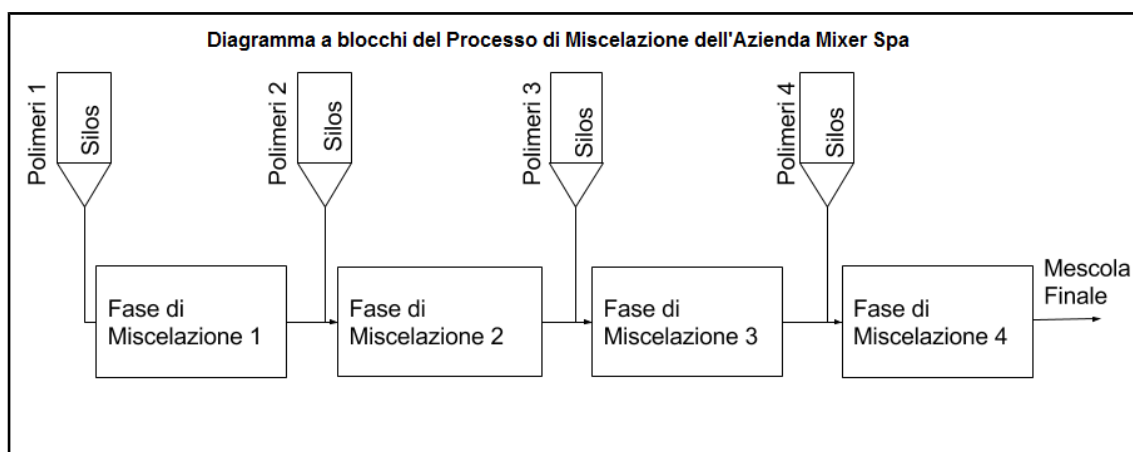
- Prova di Durezza: prova che determina la resistenza da un materiale a lasciarsi penetrare da un altro penetratore.
- Viscosimetro Mooney: strumento a disco rotante che permette di misurare la viscosità di polimeri o mescole di gomma.
- Misuratore dell'umidità: strumento che serve per la misurazione dell'umidità.
- Prova di Densità: prova per misurare il grado di compattazione del materiale.
- Microscopio Stereoscopico: strumento che viene utilizzato per studiare superfici di un campione solido.
- Melt Flow Index: indice che descrive la facilità di fluire di un polimero fuso.
- Reometro: strumento per la caratterizzazione delle mescole in gomma.

Gli output finali della linea produttiva sono pellets di materiale plastico stoccati in big bag, che dopo essere stati testati nel laboratorio presente nello stabilimento produttivo, sono venduti ad aziende produttrici di cavi di plastica.

2.4 Processo di Miscelazione

Lo scopo del monitoraggio di processo è quello di garantire la qualità del materiale plastico, per la produzione di un prodotto finale conforme alle specifiche. In questa tesi è stata analizzata unicamente la prima fase del processo produttivo ovvero quella di miscelazione, il cui svolgimento avviene in maniera discontinua in quattro sottofasi. All'inizio di ogni sottofase è previsto l'inserimento all'interno del miscelatore di nuovi ingredienti chimici posti

all'interno di silos industriali, fino alla quarta sottofase di miscelazione, che al termine, restituisce in output la miscela finale. I dati per ogni lotto sono campionati per quattro istanti successivi, in corrispondenza della fine di ogni sottofase. Le variabili di processo monitorate, sono: la temperatura della miscela, la durata della sottofase, la potenza necessaria per la miscelazione durante la sottofase e il peso della miscela. L'illustrazione 12 mostra schematicamente il processo descritto.



*Illustrazione 12: Diagramma a blocchi del Processo di Miscelazione di Tipo Batch
dell'Azienda Mixer Spa*

2.5 Approccio di miglioramento del processo

Per migliorare il ciclo produttivo di tipo batch, è necessario applicare in maniera sistematica un metodo di controllo di processo, per monitorare le proprietà del prodotto intermedio e finale considerate critiche, per verificarne la purezza la stabilità e la sicurezza desiderata. La situazione AS-IS dell'azienda Mixer prevede

un controllo di qualità Off-line, al termine dei processi produttivi, successivamente allo stoccaggio ovvero non è effettuato nessun controllo al termine della fase di miscelazione, non è valutata la conformità o meno del prodotto intermedio e non sono previste delle azioni correttive che permettano di intervenire simultaneamente al processo. In questo progetto di Tesi, è costruito uno strumento di monitoraggio che può essere inserito all'interno di un piano di miglioramento per l'azienda Mixer, che consiste nell'implementazione lungo il processo produttivo di un sistema di controllo basato su tecniche statistiche multivariate, che permetta di monitorare il processo online e di intervenire sul prodotto non conforme, evitando di far proseguire nella produzione un lotto anomalo comportando uno spreco di tempo e di costi.

L'approccio proposto prevede la costruzione di un modello matematico e di carte di controllo multivariate, utilizzando dati rappresentativi di lotti storici in condizioni operative normali.

2.6 Dati disponibili

I dati disponibili per lo studio sono dati relativi a quattro variabili delle caratteristiche del processo di miscelazione. Le variabili sono riportati nella tabella seguente. Il processo di miscelazione è suddiviso in quattro sottofasce, al termine di ogni sotto-fase è previsto un campionamento dei dati. Gli intervalli di tempo tra un campionamento e l'altro sono molto brevi, dell'ordine dei secondi. Al termine di ogni sottofase, sono misurati tramite dei sensori posti sul macchinario di miscelazione, i valori corrispondenti alle variabili presenti nella tabella 1; al termine del processo di miscelazione per lotto, si hanno a disposizione sedici misurazioni ovvero quattro valori dei parametri di processo

misurati per quattro istanti di campionamento.

Variabile del Processo Miscelazione	Tipo di Variabile
1	Temperatura della Mescola (K)
2	Durata della Sotto-fase (s)
3	Potenza (W)
4	Peso della Mescola (gr)

Tabella 1: Variabili del Processo di Miscelazione misurate dai sensori

I dati storici messi a disposizione dall'azienda Mixer riguardano 1987 lotti denominati NOC, ovvero lotti caratterizzati da condizioni operative normali e 150 lotti denominati AOC, ovvero lotti caratterizzati da condizioni operative anormali. I grafici seguenti mostrano il profilo temporale medio delle variabili di processo nei quattro istanti di campionamento. La linea nera è il valore della media aritmetica della variabile di processo e le linee tratteggiate individuano il valore medio più o meno due volte la deviazione standard. I valori medi e la deviazione standard dei parametri di processo sono calcolati utilizzando un set di lotti caratterizzati da condizioni operative normali chiamato dataset “*Training*”, su cui è costruito il modello di monitoraggio che sarà descritto nel capitolo 3.

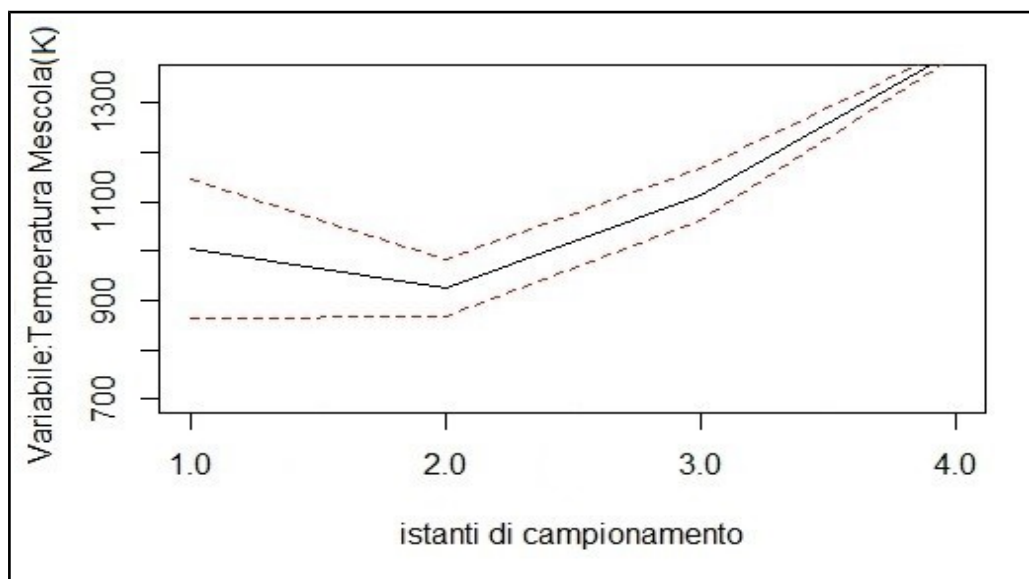


Illustrazione 13: Profilo temporale medio della Variabile della Temperatura della Mescola nei 4 istanti di campionamento

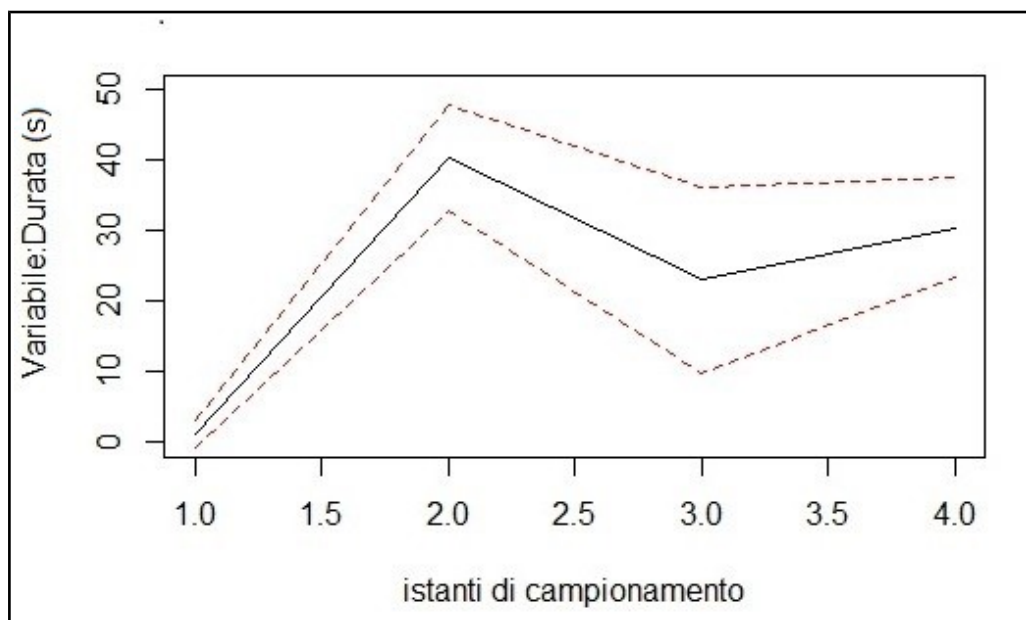


Illustrazione 14: Profilo temporale medio della Variabile della Durata della Sottofase nei 4 istanti di campionamento

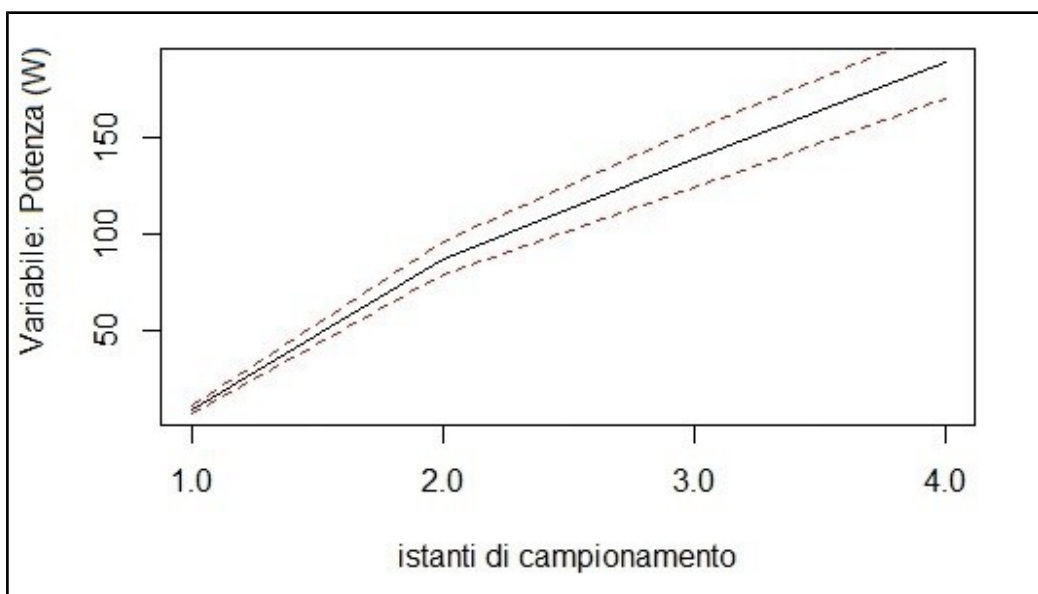


Illustrazione 15: Profilo temporale della Variabile della Potenza necessaria per la miscelazione nei 4 istanti di campionamento

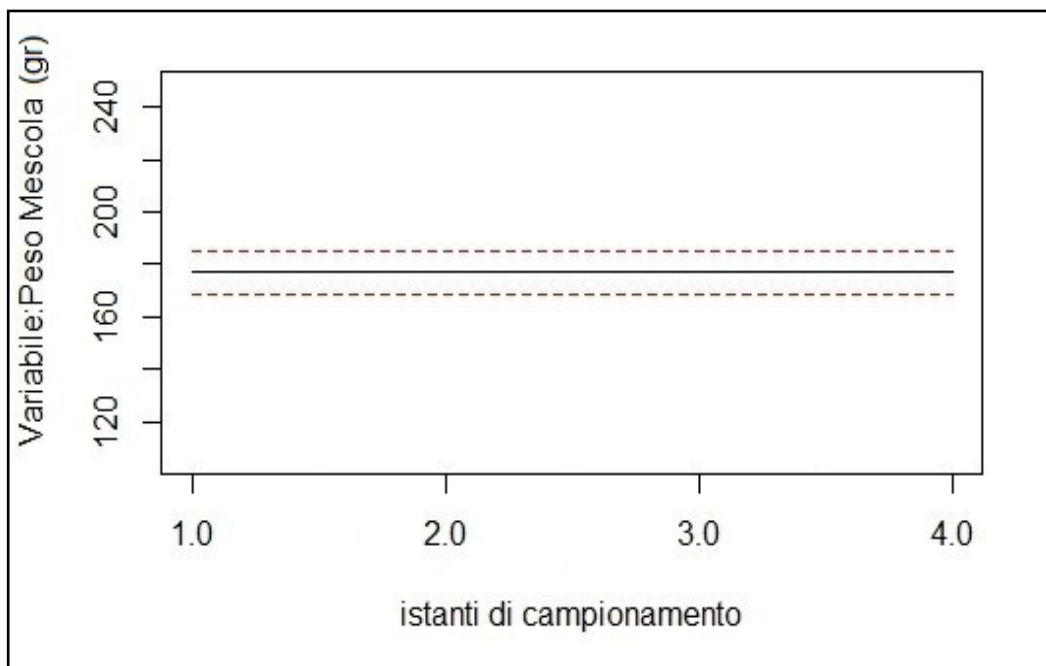


Illustrazione 16: Profilo temporale medio della Variabile del Peso della Mescola nei 4 istanti di campionamento

Capitolo 3

MSPC: Multivariate Statistical Process Control

In questo capitolo sono illustrate le tecniche statistiche multivariate, che sono state utilizzate nella tesi per il monitoraggio di processo: è descritta la metodologia per la costruzione del modello Multiway-PCA, delle carte di controllo T^2 di Hotelling, delle carte Q e dei rispettivi limiti di controllo. In ultimo è proposto un approccio alternativo per la costruzione del modello utilizzando il metodo di Tucker.

3.1 Controllo Statistico Multivariato di Processo

La qualità dei prodotti finali di un ciclo produttivo, deve essere monitorata per garantire l'assenza di scostamenti eccessivi dalle specifiche previste (superamenti), per eliminare la possibilità di danneggiamento dell'impianto, delle apparecchiature e per garantire la sicurezza del personale. I sensori, presenti lungo la linea produttiva misurano nel tempo i parametri di processo che incidono sulla qualità del prodotto. La mole di informazioni raccolte durante

lo svolgimento dei processi è molto elevata, in particolare nei processi batch e di conseguenza il controllo può trasformarsi in un esercizio ridondante e complicato. Per rendere efficiente il controllo di processo, sono state introdotte diverse tecniche di controllo statistico multivariato (Nomikos & MacGregor, 1995). Il controllo statistico di processo si basa su metodi multivariati, mediante i quali si monitorano i dati acquisiti in un impianto rispetto alle condizioni operative normali. La tecnica principalmente utilizzata per il monitoraggio statistico multivariato è l'Analisi delle Componenti Principali (MacGregor & Kourti, 1995) che può essere applicata a processi continui o a processi batch con lo sviluppo di apposite carte di monitoraggio. Con esse, fissati dei limiti di confidenza da dati riferiti a lotti "normali" di processo, si osserva, con una predeterminata soglia di probabilità, se il processo segue delle modalità di svolgimento in condizioni normali o anomale.

3.2 Carte di Controllo Univariate: Problematiche e Confronto con la PCA

Ogni processo produttivo presenta una variabilità intrinseca, che è chiamata "Variabilità Naturale," creata da un insieme di variazioni che sono interne al processo. Le cause che originano la variabilità intrinseca di processo sono dette comuni e non dipendono da elementi esterni, un processo che dipende unicamente da questo tipo di cause è definito prevedibile ed è descrivibile tramite leggi statistiche (Haridy & Wu, 2009). L'intervento di fattori esterni può determinare un cambiamento della variabilità comune del processo, creando un disturbo che incide negativamente sul corretto funzionamento del processo. I fattori esterni sono chiamati cause speciali generano una variabilità straordinaria che non è prevedibile e che può portare il processo ad essere

definito “fuori controllo”. E' necessario sapere se la variabilità è generata da cause comuni oppure da cause straordinarie. Le carte di controllo, che rappresentano un processo nel tempo, consentono di stabilire se un processo è fuori controllo o meno. Esistono in letteratura diversi tipi di carte di controllo. Nell' illustrazione 17, è mostrato un esempio generico di carta di controllo univariata introdotta Shewhart nel 1931.

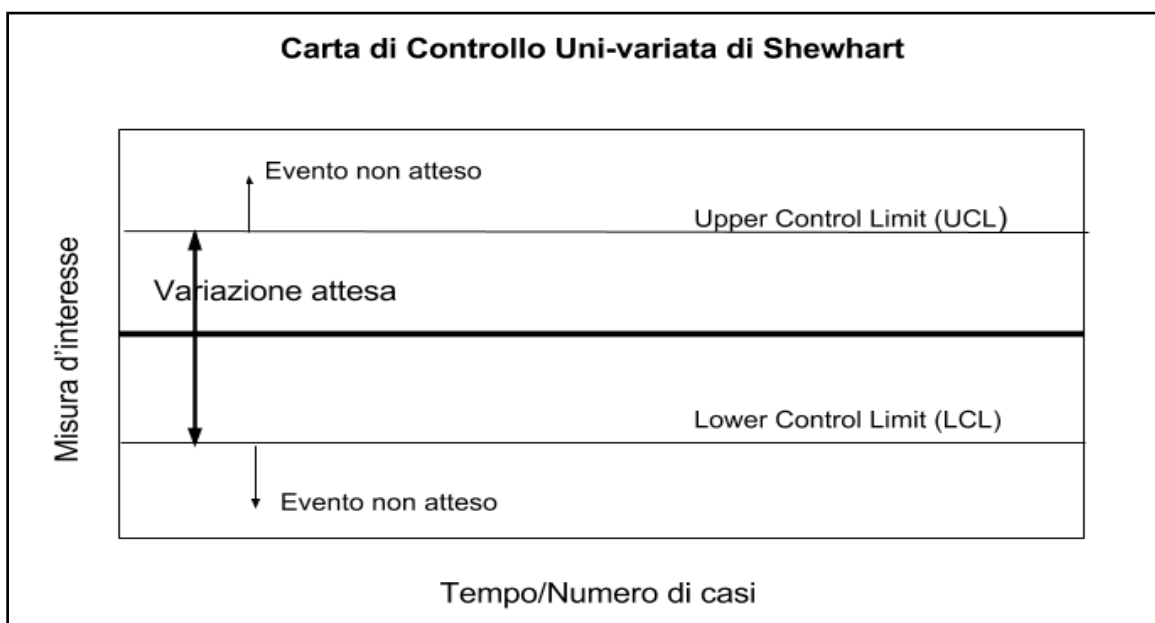


Illustrazione 17: Carta di controllo univariata di Shewhart

I valori limite per il controllo statistico della variabile x sono calcolati con le seguenti formule, rappresentative dei limiti di controllo, della deviazione standard e della media:

$$UCL = \bar{x} \pm 2s$$

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N x_i - \bar{x}}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{i1} = 0$$

Il campione che viene analizzato segue una distribuzione di tipo gaussiana di media \bar{x} e di deviazione standard s . Esiste una probabilità pari al 95% che il campione cada all'interno dei limiti di controllo. Le due linee UCL e LCL nell'illustrazione 17, sono rispettivamente il limite superiore e il limite inferiore delle carte di controllo univariate. I due limiti definiscono il range entro il quale il valore del parametro deve essere contenuto per essere definito conforme, laddove esista unicamente variabilità comune dei dati. L'Upper Control Limit e Lower Control Limit sono calcolati in funzione di una distribuzione gaussiana. L'interpretazione delle carte di controllo univariate è simile a quella di un test statistico classico. I valori sopra del limite, indicano un andamento non omogeneo del processo invece quelli che rimangono all'interno dei limiti, sono testimonianza della variabilità comune del processo. I punti della carta devono essere all'interno dei limiti ma questo non è abbastanza per asserire che un processo è in controllo. Una situazione anomala, è individuata, anche quando è riscontrabile un andamento ciclico dei campioni all'interno dei limiti o quando è presente un trend crescente, che può portare il processo fuori controllo. Le problematiche delle carte di controllo di Shewhart, si riscontrano quando le variabili sono correlate e quando il loro numero è molto elevato. L'utilizzo di carte statistiche univariate non è adatto all'analisi di processi industriali che presentano molte variabili. In tale processi la variabilità straordinaria impatta sull'equilibrio di tutte le variabili di processo, che ne determinano la qualità del prodotto finale. L'esempio seguente spiega quali sono le problematiche della carte univariate riportando il confronto tra due approcci: il primo, univariato, monitora singolarmente due parametri di processo X_1 e X_2 , e il secondo, multivariato, invece rappresenta le variabili congiuntamente.

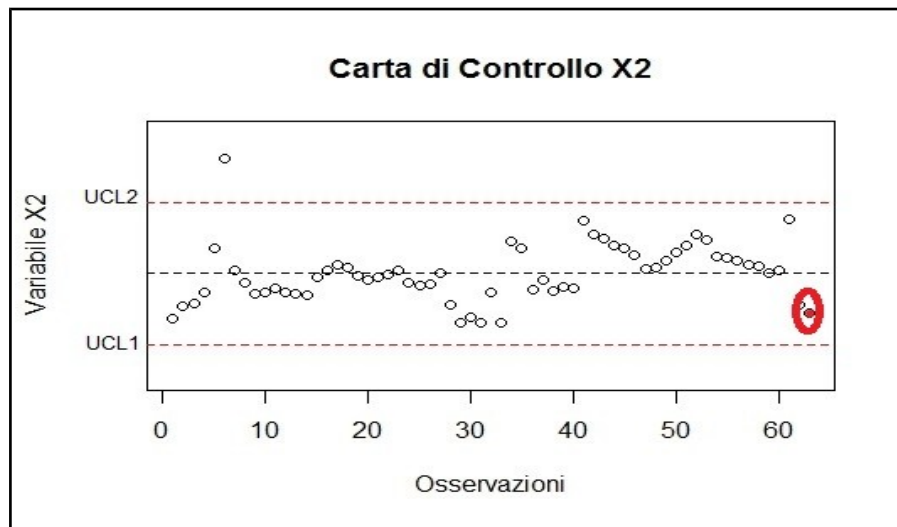


Illustrazione 18: Carte di controllo univariata per un parametro di processo X1 con campione rosso rappresentativo del lotto non conforme all'interno dei limiti di confidenza rappresentati dalle linee rosse tratteggiate

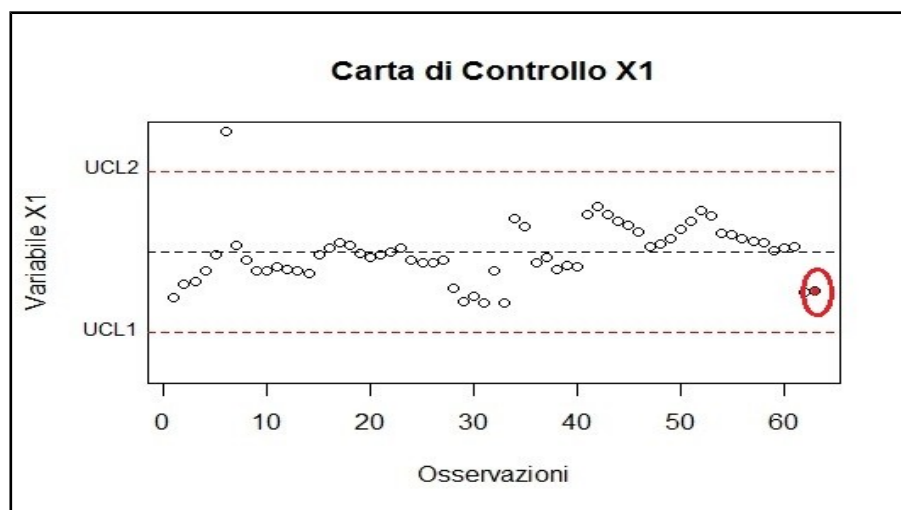


Illustrazione 19: Carte di controllo univariata per un parametro di processo X2 con campione rosso rappresentativo del lotto non conforme all'interno dei limiti di confidenza rappresentati dalle linee rosse tratteggiate

Nelle illustrazioni precedenti, sono presenti due carte di controllo univariate che monitorano due variabili di processo diverse X_1 e X_2 . Nonostante il campione rosso si trovi all'interno dei limiti di controllo, si riscontrano dei problemi, ovvero l'ultimo lotto delle carte non è conforme alle specifiche. Per spiegare questo fenomeno, il primo passo da compiere è calcolare la correlazione tra le due variabili con le formule seguenti (MacGregor & Kourti, 1995):

$$s_{12} = \frac{1}{n-1} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)$$

$$\rho_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}$$

Se la correlazione tra le variabili è molto elevata, può presentarsi il caso che è descritto successivamente. La rappresentazione congiunta delle due variabili X_1 e X_2 presente nell'illustrazione 20, permette di evidenziare come il campione di colore rosso si discosti dagli altri campioni. Le variabili X_1 e X_2 sono molto correlate, per questo motivo la maggior parte dei campioni si dispongono sulla retta bisettrice del piano cartesiano, tranne il campione 61 che si allontana dalla retta della bisettrice. Tale informazione non era visibile nei grafici delle carte univariate di controllo, dove il campione 61 veniva classificato a norma perché all'interno dei limiti di controllo. In questo esempio, è evidente come la correlazione tra solo due variabili di processo sia un'informazione necessaria da tenere in considerazione, per ottenere dei risultati rappresentativi della variabilità congiunta dei dati. Nel momento in cui il numero di variabili raggiunge un valore molto elevato, l'efficacia delle carte univariate e la loro capacità di diagnosticare un campione realmente fuori norma diminuisce esponenzialmente.

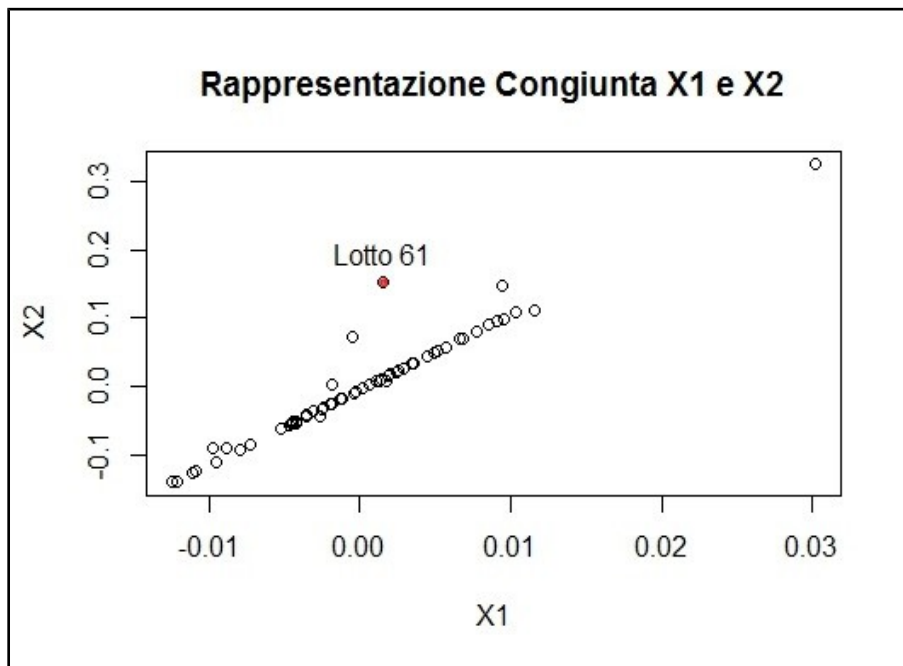


Illustrazione 20: Rappresentazione Congiunta di due Variabili di Processo

Nell'ambito di processi con un elevato numero di caratteristiche, è probabile che la causa che faccia variare l'andamento nel tempo di una sola variabile incida su tutte le altre per via delle relazioni che sono presenti tra di esse. Perciò è stato preso in considerazione un metodo multivariato, che descrive il rapporto tra le diverse caratteristiche e permette di costruire un modello di controllo da applicare simultaneamente a tutti i parametri di processo.

Per ovviare alle problematiche delle carte di controllo univariate, è quindi proposta la seguente strategia di controllo multivariato: è stato utilizzato il metodo delle componenti principali PCA e le due carte di controllo per la statistica di Hotelling T^2 e quella per la statistica Q .

3.3 PCA: Analisi delle Componenti Principali

L'obiettivo principale dell'Analisi delle Componenti Principali è la riduzione della dimensionalità di un dataset, conservando l'informazione utile e “scartando” il rumore. In questo modo si intende diminuire la dimensionalità dei dati di processi, in modo da ottenere una rappresentazione che conservi la struttura delle variabili (Gunther, Conner, & Seborg, 2007). La PCA cerca di risolvere il seguente problema, la rappresentazione di un insieme di punti in maniera bidimensionale nelle direzione che massimizzi la varianza dei dati. Prima di utilizzare la PCA , è applicato il ridimensionamento dei dati. Le traiettorie medie per ogni variabile di parametro di processo sono rimosse (Kourti, 2005). Si ipotizzi di avere dei campioni sperimentali di un vettore gaussiano $X_1.....X_n$. Questi dati si dispongono in maniera ellissoidale. Per conoscere gli ellissoidi è necessario calcolare la matrice di Covarianza S, la cui formula è la seguente.

$$S = \frac{1}{(n-1)} X^T X$$

Successivamente è necessario calcolare gli Autovalori $\lambda_1.....\lambda_n$ e gli Autovettori. Gli ellissoidi presentano come assi, gli Autovettori $P_1.....P_n$ e come lunghezza dei vettori le radici degli autovalori. Nella PCA, si effettua una trasformazione dalla base canonica dei dati originali $u_1.....u_n$ e la base delle componenti principali $e_1.....e_n$, autovettori di S, con autovalori λ . Gli autovettori $e_1.....e_n$ sono combinazioni lineari dei vettori $u_1.....u_n$. La matrice degli autovettori è detta dei “Loadings” . La formula che segue descrive la relazione che esiste tra la base canonica dei dati originali $u_1.....u_n$ e le componenti principali.

$$u_p = e_1^1 e_1 + \dots + e_p^p e_p$$

Il numero e_i^j è il coefficiente che lega e_i con u_j . I "Loadings" quantificano il legame che esiste tra le diverse variabili. I vantaggi, associati ai "Loadings" sono la creazione di un legame tra le vecchie variabili u_1, \dots, u_n e le nuove variabili e_1, \dots, e_n e lo sviluppo di modelli di legami tra le variabili (Reinikainen & Höskuldsson, 2007).

Al fine di avere la visualizzazione migliore dei punti sperimentali, gli assi che sono da considerare solo quelli di lunghezza maggiore. Per avere una visualizzazione tridimensionale dei dati si considerano i primi 3 autovettori P_1, P_2, P_3 . La matrice P presenta come colonne gli autovettori (detti "Loadings") della matrice di covarianza S , la formula è la seguente:

$$S P = P \lambda$$

Ogni autovalore è associato ad un autovettore. Il primo autovettore associato all'autovalore più grande identifica la prima Componente Principale. L'autovalore è rappresentativo della varianza spiegata della corrispondente Componente Principale. Al fine di ridurre la dimensionalità, le componenti associate agli autovalori di valore minimo possono essere eliminate, perciò scelto un numero pari ad A Componenti Principali per la creazione del modello, il set di dati finale risulterà A -dimensionale. La proiezione della matrice T (chiamati *Score*) nel nuovo spazio, definito dalle componenti principali, è la seguente:

$$T = X P$$

La proiezione T sullo spazio K-dimensionale è indicata dalla seguente formula:

$$\hat{X} = TP^T$$

dove la differenza tra X e \hat{X} è uguale alla matrice dei Residui E .

$$E = X - \hat{X}$$

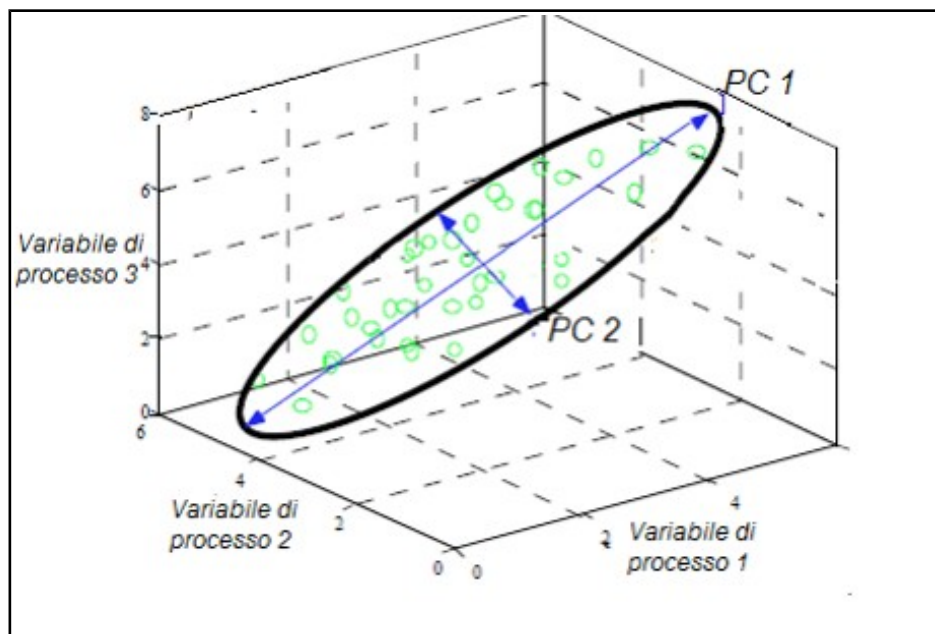


Illustrazione 21: Piano definito dalle componenti principali PC1 e PC2

La matrice degli *Score* T che ha dimensioni $I \times A$, dove I sono il numero dei lotti e A il numero delle Componenti Principali, rappresentano le coordinate dei dati originali nello spazio individuato dalle due componenti principali. Gli *Score* contengono le informazioni su come i dati sono correlati tra di loro. La matrice P dei *Loadings* ha coordinate $A \times J$, dove J sono il numero di colonne della matrice

dei dati originali. Gli *Score* sono tra loro ortogonali, invece i Loadings sono ortonormali. I Residui E identificano le informazioni, che il Modello costruito sulle Componenti Principali, non ha rappresentato (detto *Rumore*).

Per sorvegliare un processo la cui dimensionalità è stata ridotta tramite PCA, si possono utilizzare due tipologie di carte di controllo, la T^2 e la Q. La Statistica T^2 è la somma della matrice degli *Score* normalizzati al quadrato, la statistica Q è la somma dei quadrati dei residui ed è una misura della quantità della variazione non spiegata dal modello PCA.

3.3.1 Scelta del numero di componenti per il modello

La rappresentatività del modello è indicata dal valore dei residui, che deve essere minimizzata. La selezione del numero delle componenti principali è un passo fondamentale per ottenere un modello rappresentativo della variabilità la variabilità comune dei dati. I dataset multivariati contengono molte informazioni, che spesso risultano essere ridondanti, per via della correlazione che è presente tra le varie variabili. Spesso contengono a loro interno delle informazioni che sono del tutto irrilevanti che vengono definite "Rumore". La PCA è un metodo di riduzione dimensionale, che spesso accumula nelle primi componenti principali la massima variabilità. Le prime componenti hanno dei vantaggi, come quelli di facile interpretabilità all'interno del problema, ma la problematica fondamentale è la scelta del numero di componenti principali da utilizzare per la riduzione del problema.

Il metodo più diffuso per la selezione delle componenti principali, è quello che è stato proposto da Cattell nel 1966, è quello di tracciare gli autovalori in ordine di grandezza e poi cercare nel grafico, quello che viene definito "gomito" ((Auer, 2007)). Questo metodo viene chiamato dello "Scree Plot" che consiste nella

rappresentazione grafica degli autovalori. Gli autovalori, ognuno rappresentativo della varianza spiegata da ciascuna componente principale, sono riportati in maniera decrescente nel grafico, creando una linea spezzata. L'esplorazione di quest'ultimo permette di individuare una brusca variazione di pendenza, che permette di indicare qual'è il valore da scegliere, per il numero di componenti principali. Tale punto divide quelli che sono definiti, i fattori importanti da quelli meno importanti, ovvero le componenti principali da scegliere e quelle da eliminare. Questo metodo, è stato criticato per l'interpretazione soggettiva, perché non vi è una regola oggettiva che suddivide quelli che sono i fattori importanti da quelli irrilevanti.

Un'altra metodologia utilizzata, applicata nella tesi, è quella di raggiungere una certa quota di varianza spiegata, di conseguenza scegliere il numero di componenti principali. Un altro metodo, è quello di valutare l'introduzione di una nuova componente principale e individuare l'aumento della varianza spiegata di tutte le componenti con l'aggiunta della nuova, se l'aumento è significativo la componente può essere inserita nel modello.

3.4 Costruzione del modello e delle Carte di controllo multivariate

Da un dataset di lotti storici detti *Golden Batches (lotti d'oro)*, che costituiscono il set chiamato di *Training*, perché composto da lotti caratterizzati da normali condizioni operative, si costruisce il *Modello* di legami tra le variabili di processo, ovvero si determinano i *Loadings* e la matrice di Covarianza avente sulla diagonale gli autovalori.

Costruito il modello sul set di dati *Training*, si possono proiettare sullo spazio definito dalle componenti principali (nuovi assi dello spazio ridotto dalla PCA), i dati di un Set di lotti *Test*, per verificare se presentino la stessa variabilità dei dati

del dataset di *Training* (Laursen, Frederiksen, Leuenhagen, & Bro, 2010). Nella formula seguente è presente la proiezione dei dati *Test* X nello spazio delle componenti principali.

$$T = X P$$

La proiezione dei dati, considerando A componenti principali, è calcolabile con la formula seguente:

$$t_A = X P_A$$

Per individuare i campioni fuori norma sono utilizzate le statistiche di controllo T^2 e Q . Per stabilire se i nuovi dati proiettati sulla nuova base siano in condizione di normalità, si utilizzano i limiti delle due statistiche calibrandole sul Modello.

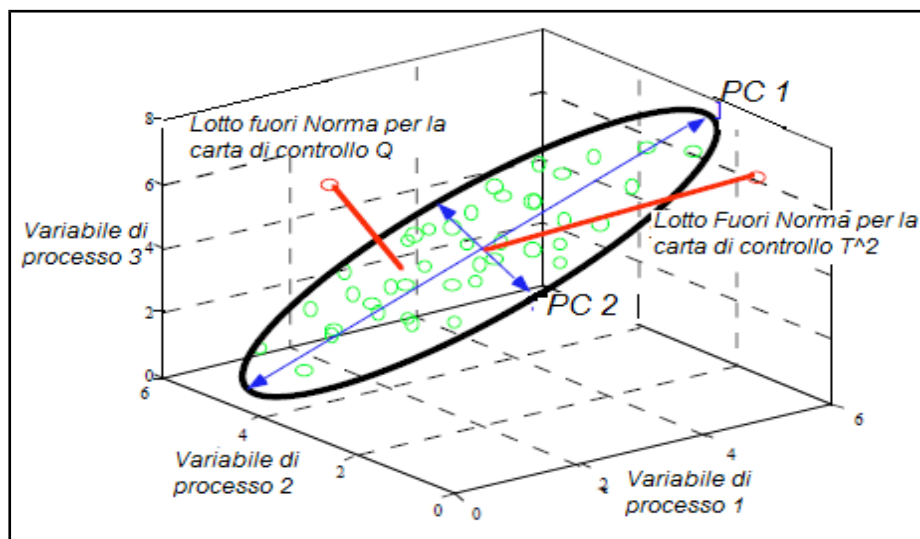
La statistica Q misura la distanza del campione i -esimo dall'iperpiano di dimensioni ridotte, costituito dalle Componenti Principali e la statistica di Hotelling T^2 misura la distanza di un campione i -esimo dall'origine del modello PCA. In particolare la statistica T^2 è uguale alla somma degli "Score" normalizzati e la statistica Q è la somma dei residui al quadrato. Le formule per il calcolo delle statistiche T^2 e Q , in riferimento all' i -esimo campione (Westerhuis, Gurden, & Smilde, 2000), prendono in considerazione le prime A componenti principali. Di seguito sono riportate le formule di entrambe le carte di controllo.

$$T_i^2 = \sum_{j=1}^A \frac{t_{ji}^2}{\lambda_j}$$

$$Q_i = \sum_{j=1}^J (x_{ij} - \bar{x}_{ij})^2$$

I limiti di controllo delle carte di controllo sono detti T^2_{limite} e Q_{limite} , la cui costruzione è descritta nel paragrafo 3.6 (Gunther et al., 2007).

Se il lotto i-esimo presenta il valore T^2 superiore al T^2_{limite} e il valore Q superiore del Q_{limite} , è definito fuori norma.



*Illustrazione 22: Piano definito dalle componenti principali
PC1 e PC2 circoscritto dai limiti di controllo T^2 e
rappresentazione di lotti fuori norma per entrambe le carte
di controllo (campioni in rosso)(Wise et al., 1997)*

L'illustrazione 21 mostra il piano definito dalle componenti principali, circoscritto dai limiti definiti dal T^2 e Q (Wise, Gallagher, Butler, White, & Barna, 1997). I dati si trovano principalmente su un piano descritto da due Componenti Principali del Modello. Il primo autovettore identifica la massima variabilità dei dati (PC1, prima componente principale) ed è ortogonale al secondo autovettore (PC2, seconda componente principale). Due campioni non si adattano al modello: un lotto supera il limite di controllo di T^2 , che crea un ellisse sul

Piano Principale, un altro lotto supera il limite di controllo delle statistica Q. La misura del test Q identifica quanto il lotto campione è conforme al modello.

3.5 Trattamento preliminare dei dati

Al fine di estrarre le caratteristiche di correlazione la matrice dei dati originali deve essere pretrattata, in questo caso di studio è stato utilizzato l'autoscaling. L'autoscaling consiste in un centramento sul valore medio e in una riduzione a varianza unitaria del dataset originario dei parametri di processo (Gunther et al., 2007). Il centramento si ottiene sottraendo la media di ogni variabile del modello ai nuovi dati. Lo scaling compensa le differenze di unità di misura delle variabili e la differenza di "intensità" numerica, in modo da dare a tutte lo stesso peso. Si effettua dividendo tutte le misure di una variabile per la deviazione standard della variabile stessa calcolata in base al dataset *Training*.

3.6 Carta di Controllo Riassuntiva

Una carta di controllo riassuntiva di immediata lettura è visibile nell'illustrazione 23. La carta riassuntiva unisce i risultati della carta di controllo T^2 e Q. La NOR (Normal Operation Region) è la regione rettangolare della carta, dove ricadono i campioni a norma (ovvero i lotti conformi). I valori di T_i^2, Q_i sono le due statistiche calcolate per ogni i-esimo lotto. Le rette tratteggiate in rosso sono i limiti delle Carte di Controllo T_{limite}^2 e Q_{limite} .

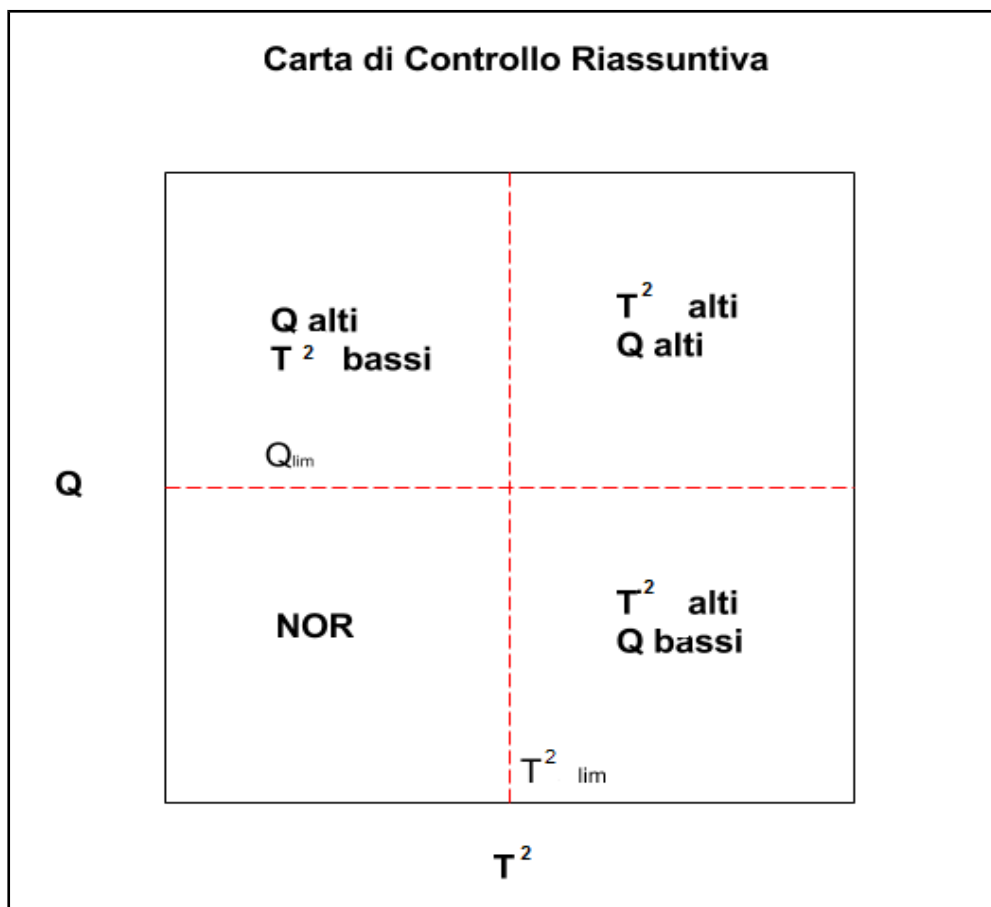


Illustrazione 23: Carta Riassuntiva delle carte di controllo T^2 e Q

Questa tipologia di Carta , incrociando i dati di entrambe le carte di controllo T^2 e Q , dà un immediata lettura dei campioni che risultano fuori norma che risultano essere al di fuori della regione NOR. In particolare sono classificati come anomali, i campioni che presentano sia T^2 che Q oltre i limiti di controllo.

3.7 Multiway-PCA

L'approccio Three-way della Multiway Principal Component Analysis crea una riduzione della dimensionalità di un dataset tridimensionale. La MPCA è equivalente all'esecuzione ordinaria PCA, su una matrice bidimensionale aperta costruita dall' "Unfolding" della matrice di dati in tre dimensioni, dove una dimensione sono i lotti (dimensione I), l'altra dimensioni le variabili di processo (dimensione J) e la terza dimensione il tempo (dimensione K) (Garcia, Ruiz, Colomer, & Melendez, 2015). Il cosiddetto "Unfolding" o Dispiegamento della matrice (Illustrazione 23) può avvenire in diversi modi (Acar, Aykut-Bingol, Bingol, Bro, & Yener, 2007). Il primo metodo è l'Unfolding nella direzione delle variabili, ogni "Slice" orizzontale di dimensione $J \times K$ (dove J sono le variabili e K gli istanti di campionamento) viene disposta verticalmente rispetto la precedente in modo da ottenere una matrice di dimensione $K \times J$, che equivale a rappresentare i dati delle variabili negli istanti di tempo campionati.

Il secondo modo è quello nella direzione dei lotti, proposto nella Tesi (Illustrazione 24). Le matrice di dimensioni $I \times J$, dove I sono i lotti e J le variabili, sono affiancate orizzontalmente in modo da ottenere una matrice di dimensioni $I \times JK$, in cui ogni riga presenta i dati di un lotto per tutte le variabili di processo, in tutti gli istanti di campionamento. Nel monitoraggio in linea di processo, il nuovo lotto x_{new} deve contenere i dati dell'intero batch ovvero i valori di tutti i campionamenti JK per ogni variabile in ogni istante temporale.

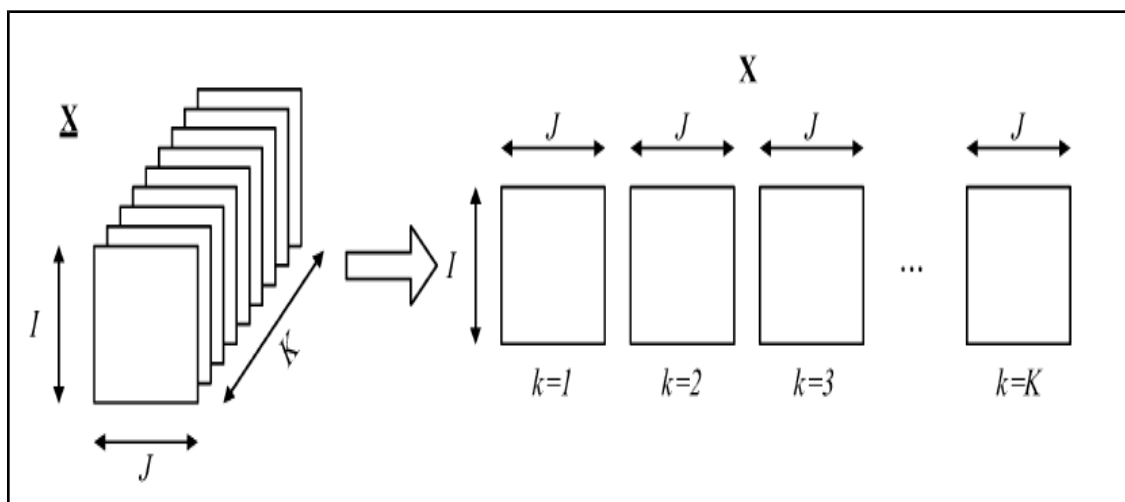


Illustrazione 24: Unfolding della matrice Three-Way nella direzione dei Batches I, in una matrice bidimensionale di dimensioni $I \times JK$ (Lotti x Variabili/Tempo) (Gunther et al., 2007)

Il processo precedente permette di decomporre la matrice tridimensionale dei dati X in una matrice bidimensionale, ottenuta dal prodotto di Kronecker tra la Matrice degli *Score* dei nuovi dati campione X (ottenuti dalla proiezione dei dati di processo originari sul modello MPCA come nella PCA) e la matrice dei *Loadings* del modello MPCA, sommato alla matrice E dei Residui, di seguito è riportata la formula (Ruiz, Colomer, & Mel, n.d.).

$$X = T \otimes P + E$$

Nell'illustrazione 25, è possibile vedere la scomposizione che la Multiway PCA crea su una matrice tridimensionale, di dimensioni I (Lotti), J (Variabili) e K (istanti di campionamento), considerando A componenti principali per la costruzione del modello.

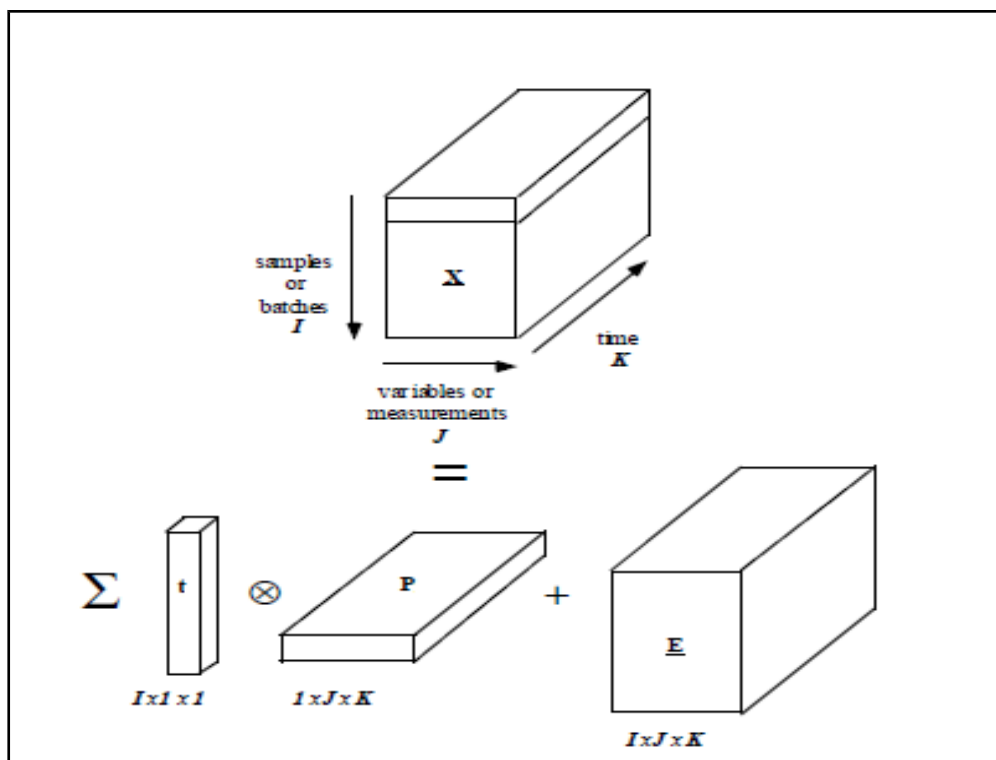


Illustrazione 25: Array di dati Tridimensionali e Decomposizione

Multyway PCA (Wise et al., 1997)

Anche per il monitoraggio di processo utilizzando la Multiway-PCA, sono necessarie le due carte di controllo, la statistica Q e la statistica T^2 . Il nuovo lotto viene confrontato con i limiti di controllo per determinare se il processo è in controllo o meno. I limiti di controllo sono calcolati selezionando solo i lotti caratterizzati da normali condizioni operative (NOC) su cui è costruito il modello, in altre parole qualsiasi situazione anomala deve essere rimossa in questa fase, da questo dipende l'applicazione di successo della carta di controllo.

Le formule per il calcolo della statistica Q e per il calcolo della statistica T^2 , nel caso della Multiway PCA si differenziano leggermente dalla PCA a due vie, sono indicate successivamente (Garcia et al., 2015). La statistica Q per l' i -esimo lotto, per j variabili, per k istanti temporali è calcolata come:

$$Q_i = \sum_{j=1}^J \sum_{k=1}^K (e_{ijk})^2$$

dove e_{ijk} sono gli elementi di E per i-esimo lotto. La statistica T^2 per l' i-esimo lotto, per j variabili, per k istanti temporali è calcolata come:

$$T^2 = t_i^T S^{-1} t_i$$

dove t_i sono gli Score dell' i-esimo lotto ed S è la Matrice di Covarianza. Il nuovo lotto viene confrontato con i limiti di controllo per determinare se il processo è in-control oppure out-control.

3.8 Limiti di Controllo delle Carte

Dopo aver calcolato il modello è necessario calcolare i limiti di confidenza del sistema, ovvero i limiti di controllo che il modello ha nella rappresentazione di nuovi campioni. I limiti di controllo sono calcolati in base al numero di lotti presenti nel dataset *Training*, su cui è stato costruito il modello Multiway-PCA. Gli *Score* sono combinazione lineari delle variabili originali, si suppone che essi possano essere approssimati ad una distribuzione normale. Si assume che le coordinate dei nuovi dati seguano una distribuzione F di A (componenti principali), con I-A gradi di libertà, dove I sono i lotti su cui è stato costruito il modello (Gunther et al., 2007). L' Upper Control Limit per il test statistico T^2 al livello di significatività α , è dato dalla seguente formula (Zhao, Wang, Mao, Lu, & Jia, 2008) :

$$UCL(T_{limite}^2)_\alpha = \left(A \frac{I^2 - 1}{I(I - A)} \right) F_{(A, I-A), \alpha}$$

Il limite di confidenza di T^2 , calcolato con la formula precedente, definisce un ellisse sul piano nel quale è proiettato il lotto in condizioni operative normali.

Per calcolare Upper Control Limit (UPC) del test statistico Q, esistono diverse procedure, in questa tesi è stata applicata la procedura di Jackson e Mudholkar. I limiti di controllo sono stati fissati ad un livello di significatività pari al 5% (Gunther et al., 2007).

$$UCL(Q_{limite})_\alpha = \theta_1 \left[1 + \frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{(\frac{1}{h_0})}$$

dove $V = \frac{E E^T}{I - 1}$ è la matrice di covarianza ed E è la matrice dei residui $I \times JK$,

ovvero la matrice resa bidimensionale tramite “Unforlding”.

La formula seguente presenta il calcolo di θ_i , ovvero la traccia della matrice di covarianza V elevata ad i .

$$\theta_i = \text{trace}(V^i)$$

Attraverso i valori degli omega, è possibile calcolare il valore di h_0 , necessario per il calcolo del limite di controllo della carte Q.

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$$

3.9 Validazione del Modello e Calcolo delle Sensitività

La fase di validazione dal modello è necessaria per conoscere la percentuale di falsi allarmi segnalati e la percentuale di veri allarmi non segnalati dalle carte di controllo, costruite in base al modello MPCA. Siano definiti NOC, i lotti che presentano Condizioni Operative Normali e AOC, i lotti che presentano Condizioni Operative Anormali.

L'approccio sperimentale che è stato utilizzato prevede di utilizzare 3 dataset:

a) **Dataset Training:**

Il dataset di *Training*, come è stato definito precedentemente, è costituito da lotti NOC (lotti in condizioni operative normali) ed è utilizzato per costruire il modello Multiway-PCA, che descrive la variabilità del processo in condizioni operative normali.

b) **Dataset Test NOC:**

Il secondo dataset è il Set-Validazione NOC, costituito da lotti NOC in normali condizioni operative, lotti diversi da quelli utilizzati per la costruzione del modello Multiway-PCA. Questo dataset è costruito per testare la capacità delle carte di controllo costruite in base al modello MPCA, di diagnosi di lotti in condizioni di normale funzionamento. Dai dati reali del dataset Test NOC sono calcolati i rispettivi *Score* sul modello MPCA e i rispettivi valori delle due carte di controllo che sono confrontati con i limiti di controllo di Q e T^2 . Il rapporto tra i lotti test NOC che non superano limiti di controllo e il totale di lotti test NOC è la Sensitività NOC del modello, parametro che indica la capacità di diagnosi

del modello di lotti in condizioni operative normali (Garcia et al., 2015). La formula è la seguente:

$$SENSITIVITA\ NOC = \text{lotti test NOC sotto il limite} / \text{lotti test NOC totali}$$

c) **Dataset Test AOC:**

Il secondo dataset è il Set-Validazione AOC, costituito da lotti AOC in Anormali Condizioni Operative. Questo dataset è costruito per testare la capacità delle carte di controllo costruite in base al modello MPCA, di diagnosi di reali situazioni anomale associate ai lotti. Dal dataset di lotti Test AOC sono stati calcolati i rispettivi *Score* sul modello MPCA e i valori delle carte di controllo che sono confrontati con i limiti di controllo di Q e T^2 . Il rapporto tra i lotti test AOC che superano questi limiti e il totale di lotti test AOC è la Sensitività AOC del modello, parametro che indica la capacità di diagnosi del modello di lotti in condizioni operative anormali (Garcia et al., 2015). La formula è la seguente:

$$SENSITIVITA\ AOC = \text{lotti test AOC sopra limite} / \text{lotti test AOC totali}$$

3.10 Analisi Three Way dei dati: Modello Tucker 3

I dati three-way dei lotti possono essere analizzati con vari modelli matematici. Nei parametri precedenti è stato descritto il metodo della Multiway-PCA, di seguito è presentato un altro approccio: il modello Tucker 3.

Il modello Tucker 3 è un metodo più complesso rispetto alla multiway-PCA. Il vantaggio di Tucker 3 è quello di offrire una migliore riduzione dei dati originali,

poiché sono compressi in 3 direzioni, ognuna corrispondente ad un modello diverso, caratterizzato dalle proprie componenti (Van sprang et al., 2005).

Lo scopo dell'applicazione del Modello Tucker3 :

- Costruire tramite l'approccio Tucker3, partendo da dati tridimensionali di lotti storici, un modello che descriva la normale variazione nel tempo dei parametri di processo.
- Costruire le carte di controllo e limiti di confidenza statistici sulla base del modello.
- Proiettare sul modello nuovi dati batch e validare la robustezza delle carte.

3.10.1 Modello Three Way Tucker3

I dati batch three-way presentano 3 dimensioni $I \times J \times K$, con I lotti, J variabili e K istanti di tempo. La matrice dei dati three-way deve essere aperta in due dimensioni, questa fase viene chiamata di *“Unfolding”* esattamente come nella Multiway-PCA (Booksh, Bronk, & Czege, 2009). A ognuna delle 3 dimensioni di una matrice three-way viene assegnato un modello, alla dimensione I viene associato il modello *“Batch”*, alla dimensione J viene associato il modello *“Variables”* e alla dimensione K viene associato il modello *“Time”*. Per effettuare l' *“Unfolding”* della matrice è possibile mantenere il primo modello detto *“Batches”* e unire il modello delle variabili e del tempo in una nuova dimensione *“Variables_Time”* (Unfolding è spiegato nel paragrafo 3.7). Il modello Tucker 3 è utilizzato per la decomposizione e la compressione di dati di processo Three-way. Tucker 3 è un modello multilineare, che sintetizza le informazioni presenti in una matrice Three-way X, estraendo diverse componenti per ognuno dei 3

modelli, in cui decompone i dati. Nella Multiway PCA, i dati Three-Way sono scomposti in un'unica direzione costruendo un modello con A componenti principali, in Tucker3 sono costruiti 3 modelli, ognuno con le proprie componenti ed i tre modelli sono messi in correlazione tramite una matrice chiamata "Core Matrix".

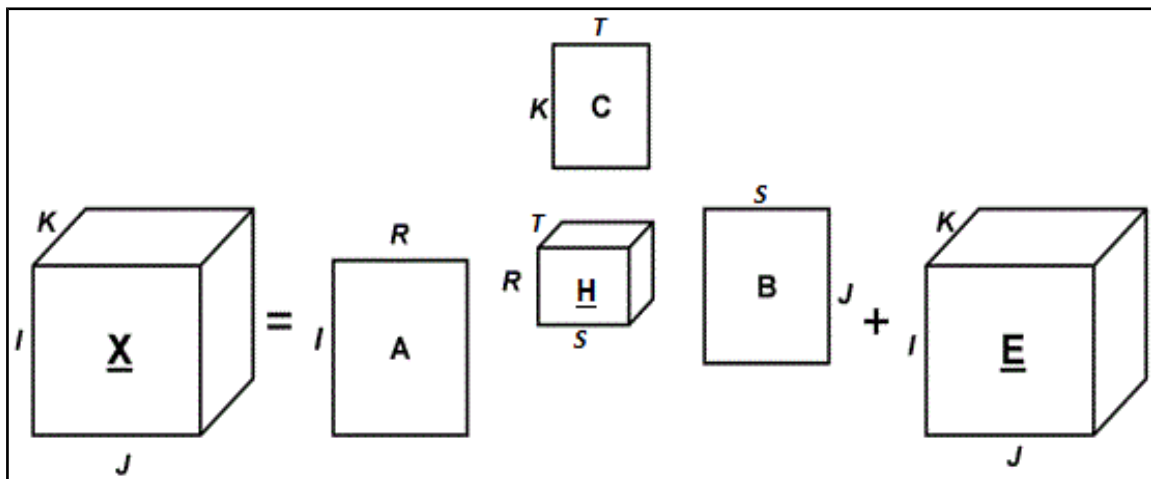


Illustrazione 26: Proiezione dei dati reali sui tre modelli di riferimento di Tucker3 modello A, modello B e modello C (Acar et al., 2007)

Il modello Tucker3 costruito per la matrice X tridimensionale $I \times J \times K$, è descritto dalla seguente formula:

$$x_{ijk} = \sum_{r=1}^R \sum_{s=1}^S \sum_{t=1}^T a_{ir} b_{js} c_{kt} h_{rst} + e_{ijk}$$

I valori R, S, T sono il numero delle componenti per i modelli A, B, C e

$x_{ijk}, a_{ir}, b_{js}, c_{kt}, h_{rst}, e_{ijk}$ sono gli elementi delle seguenti matrici:

- X ($I \times J \times K$): Matrice Three-Way dei dati di processo
- A ($I \times R$): Modello dei lotti

- B(JxS): Modello delle variabili di processo
- C(KxT): Modello degli istanti di campionamento
- H(RxSxT): Core Matrix
- E(IxJxK): Matrice dei Residui

I vettori di A sono chiamati *Score* e i vettori di B e di C sono chiamati *Loadings*. Gli elementi di H sono i pesi di tutte le possibili interazioni tra i vari modelli e la matrice H è chiamata Core Matrix. La scomposizione dei dati originali Three-way X nei 3 piani di proiezione indicati dal modello A, B, C correlati tramite la Core Matrix H, è rappresentata nell' Illustrazione 26.

Una rappresentazione bidimensionale della matrice X tramite il modello di Tucker3 è data dalla relazione seguente (Van sprang et al., 2005):

$$X = AH(C \otimes B)^T + E$$

Il valore di X è ottenuto dal prodotto tra la matrice A degli *Score* e la matrice del Core H (RxST) che è una rappresentazione bidimensionale della matrice H(RxSxT) e la matrice trasposta del prodotto di Kronecker tra la matrice dei *Loading* C e B, sommato alla matrice dei residui E. In Tucker3 il numero delle componenti può cambiare in ciascuno dei tre modelli. Per costruire i diagrammi di controllo di MSPC è necessario costruire un modello partendo dai dati lotti che rappresentano le condizioni operative normali.

Gli *Score* dei dati di nuovi lotti proiettati sul modello A di Tucker3, sono ottenuti con le formule seguenti.

$$V = (H(C \otimes B)^T)^T$$

$$t_{new} = x_{new} V (V^T V)^{-1}$$

I valori t_{new} sono gli *Score* di coordinate $I \times R$ dei nuovi dati batch, proiettati sul modello A dell'approccio Tucker3 ed x_{new} sono i nuovi dati batch di coordinate $I \times J \times K$, ovvero la matrice “*Unfolding*” dei dati dei nuovi batches. I residui del modello possono essere calcolati con la formula seguente applicabile anche per i dati di nuovi lotti.

$$E = X - AH(C \otimes B)^T$$

Dove E è la matrice bidimensionale dei Residui $I \times J \times K$, X è la matrice bidimensionale dei dati iniziali $I \times J \times K$, A ($I \times R$) è la matrice degli *Score*, H è la Core Matrix di dimensioni ($R \times S \times T$), C è la matrice dei loadings ($K \times T$) e B è la matrice dei loadings ($J \times S$).

3.10.2 MSPC Chart per il modello Tucker3

Le carte di controllo che sono state utilizzate per il modello Tucker3 si basano sulla statistica di Hotelling (carta di controllo D) e sulla statistica dei residui (carta di controllo Q) (Camacho, 2007).

$$Q_i = \sum_{j=1}^J \sum_{k=1}^K (e_{ijk})^2$$

$$D_i = \sum_{r=1}^R \left(\frac{t_{ri} - \mu_{tr}}{\sigma_{tr}} \right)^2$$

Un altro modo per scrivere la formula della carta di controllo D è presente successivamente (Van sprang et al., 2005):

$$(a_{new} - \hat{a}_I)^T S_R^{-1} (a_{new} - \hat{a}_I) I \frac{(I-R)}{R(I^2-1)}$$

I è la dimensione dei dati che rappresenta i lotti, J è la dimensione delle variabili, K è la dimensione degli istanti di tempo, e_{ijk} residui dei nuovi dati batch calcolati come differenza tra i dati originali dei nuovi lotti e i valori proiettati degli stessi sul modello three-way. R è il numero delle componenti principali del modello A dell'approccio Tucker3, t_{ri} è il vettore del r-esima componente del modello A dell'i-esimo nuovo batch, μ_{tr}, σ_{tr} sono rispettivamente la media e la deviazione standard degli *Score* dell' r-esima componente del modello A. Gli *Score* sono combinazione lineari delle variabili originali, è supposto che essi possano essere approssimati ad una distribuzione normale F di R (componenti principali) e I-R gradi di libertà. L'Upper Control Limit per il test D-statistics al livello di significatività α è dato dalla formula seguente:

$$UCL(D)_\alpha = \left(R \frac{I^2-1}{I(I-R)} \right) F_{(R, I-R), \alpha}$$

Per calcolare upper control limit (UPC) del Q-statistics esistono diverse procedure, in questa tesi è stata applicata la procedura di Jackson e Mudholkar

(Formula 3.31). I limiti di controllo sono stati fissati ad un livello di significatività pari al 95%.

$$UCL(Q)\alpha=\theta_1\left[1+\frac{z_\alpha\sqrt{2\theta_2h_0^2}}{\theta_1}+\frac{\theta_2h_0(h_0-1)}{\theta_1^2}\right]^{\left(\frac{1}{h_0}\right)}$$

I valori di θ_i sono calcolati nello stesso modo utilizzato per la MPCA.

Capitolo 4

Applicazione delle Carte di Controllo Multivariate per il monitoraggio di un processo di Miscelazione

In questo Capitolo è descritta la costruzione del modello matematico e delle Carte di Controllo per il monitoraggio del processo di miscelazione, per la produzione dei pellets in materiale plastico, i cui dati sono stati forniti dall'azienda Mixer Spa. Il modello è costruito su un dataset riguardanti lotti storici in condizioni operative normali denominato dataset di *Training*. Le prestazioni delle Carte di Controllo costruite, sono valutate tramite la proiezione sul modello di due dataset Test, il primo contenente dati di lotti NOC in condizioni operative normali e il secondo contenente dati di lotti AOC in condizioni operative anormali. In ultimo è discusso il metodo del “Contribution Plot” utilizzato per identificare le variabili di processo determinanti e responsabili della variabilità straordinaria che porta un lotto a superare i limiti di controllo.

4.1 Multivariate Statistical Process Control

Il Multivariate Statistical Process Control ha 3 obiettivi che sono il rilevamento, la localizzazione e la diagnosi di variazioni anomale di processo. Il monitoraggio del processo batch viene suddiviso in 4 fasi:

1. Fase iniziale. La prima fase consiste nella raccolta di dati di lotti storici e nella creazione dei Set di lotti "Training" e "Test".

2. Fase di training e validazione del modello. I lotti che presentano condizioni operative anomali vengono rimossi dal dataset di lotti storici "Training". Quando si raggiunge un set rappresentativo, viene costruito un modello per dimostrare la variazione comune dei dati. Dal modello ottenuto sono calcolati le statistiche T^2 , Q e i limiti di confidenza statistica per costruire i diagrammi di controllo corrispettivi (T^2 , Q). I dati "Test" sono proiettati sullo spazio definito dalle componenti principali del modello MPCA, per validare la robustezza delle Carte.

3. Fase di applicazione In questa fase il nuovo lotto viene monitorato per mezzo delle carte di controllo, derivanti dalla Fase di "Training".

4. Fase di individuazione dell'anomalia del lotto e delle variabili responsabili

Per ogni lotto fuori norma è possibile calcolare il contributo di ciascuna variabile di processo al T^2 , per determinare quale variabile e quale istante di tempo sia maggiormente responsabile della variazione straordinaria del nuovo lotto monitorato rispetto la variabilità comune del modello.

4.2 Dati di Processo

I dati storici analizzati raccolgono le misurazioni di quattro parametri di processo ottenute dai sensori posti sul macchinario di miscelazione dell'azienda Mixer. Il suddetto processo è composta da 4 sottofasi di miscelazione. Ogni parametro di processo è misurato per quattro istanti temporali successivi, al termine di ogni sottofase . Le variabili di processo misurate dai sensori sono:

- Temperatura della mescola
- Peso della mescola
- Potenza necessaria per la miscelazione
- Durata della sottofase

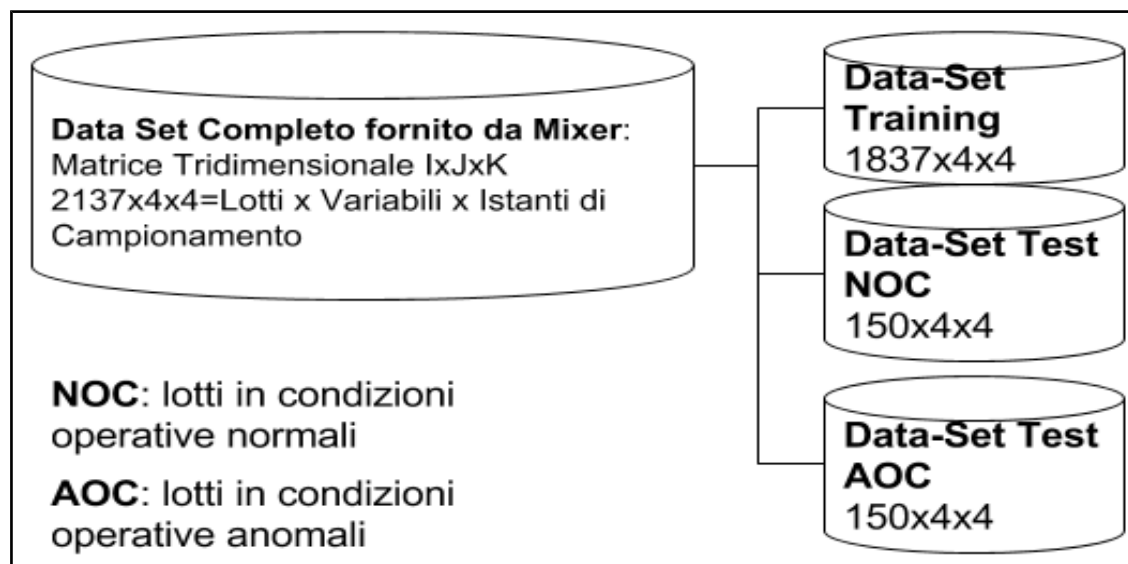


Illustrazione 27: Dataset analizzati nel caso studio: dataset “Training”, dataset “Test NOC” e dataset “Test AOC”

Sono stati creati 3 dataset dai dati storici forniti dall'azienda Mixer Spa: il primo include il Set di lotti *“Training”* necessario per la costruzione del modello MPCA e delle Carte di Controllo, il secondo e il terzo sono rispettivamente il Set *“Test”* di lotti in condizioni operative anormali (detti AOC) e il Set *“Test”* di lotti in condizioni operative normali (detti NOC), necessari per testare la validità delle Carte di Controllo (Illustrazione 27). Le matrici 3 dataset (*“Training”*, *“Test NOC”*, *“Test AOC”*) hanno una struttura tridimensionale, una dimensione I è rappresentativa dei lotti, J delle variabili di processo e K degli istanti di campionamento. In particolare il Set di lotti *“Training”* è costituito da 1837 lotti (dimensione I=1837), per ogni lotto sono presenti i valori delle misurazione dei 4 parametri di processo precedentemente elencati (dimensione J=4), campionati per 4 istanti di tempo successivi (dimensione K=4), quindi la matrice del Set di lotti *“Training”* presenta le seguenti dimensioni: 1837x4x4 (IxJxK). La matrice del Set di lotti *“Test NOC”*, formata da 150 lotti storici in condizioni operative normali ha dimensioni 150x4x4. La matrice del Set di lotti *“Test AOC”*, formata da 150 lotti storici in condizioni operative anormali ha dimensioni 150x4x4. Le matrici dei 3 dataset sono stata ridimensionate prima della costruzione del modello e delle carte, eliminando le traiettorie medie e dividendo per le deviazioni standard ciascuna colonna.

4.3 Costruzione del Modello e dei Limiti di Controllo delle Carte

Dal dataset *“Training”*, composto da 1837 lotti NOC, è costruito il modello Multiway-PCA che descrive il legame tra le variabili di processo considerate, ovvero si determinano la matrice dei *“Loadings”* e la matrice di Covarianza che ha sulla diagonale gli autovalori. Si calcola la matrice degli *“Score”* dei dati originali, ovvero le nuove coordinate dei dati proiettati sul piano descritto dalle

due prime componenti principali. In questo modo è effettuata una rotazione della nuvola multidimensionale dei dati di processo originari, per ottenerne una rappresentazione su un piano bidimensionale che ne descrive la massima variabilità. Successivamente utilizzando gli “Score” e i “Loadings” calcolati, sono determinati i limiti per entrambe le carte di controllo T_{limite}^2 e Q_{limite} , utilizzando le formule descritte nel paragrafo 3.6. Il numero di componenti principali scelte per la costruzione del modello è 3 e la varianza cumulativa spiegata dalle stesse è il 66,41%. La validità delle Carte di Controllo, ovvero la loro capacità di individuare lotti in condizioni anormali, è testata tramite due dataset come descritto in precedenza, il primo costituito da 150 lotti NOC (in condizioni operative normali) e il secondo da 150 lotti AOC (in condizioni operative anomale). I dati originari di entrambi i dataset, sono preprocessati sottraendo ad ogni colonna la media delle variabili, calcolata sul dataset “Training” (quello del modello) e dividendo ogni elemento delle colonne dei dataset “Test” per la deviazione standard calcolata sul dataset “Training” dataset “Training”. In seguito, sia per il dataset “Test NOC” e sia per il dataset “Test AOC” sono calcolati i nuovi Score. Essendo state scelte 3 componenti principali, per ogni stringa di dati di un lotto “Test” (di dimensioni reali 1x16, 1 sta per il lotto e 16=4x4 per le 4 misurazioni delle variabili di processo ripetute nei 4 istanti di campionamento successivi) si ottengono le proiezioni dei dati originari (Score) di dimensioni 1x3 (1 sta per il lotto test, e 3 per le componenti principali selezionate).

Per ogni lotto “Test” i-esimo sono calcolati i valori di T_i^2 e Q_i e confrontati con i valori dei limiti di controllo T_{limite}^2 e Q_{limite} costruiti in base al Set di lotti “Training”. Il lotto i-esimo sottoposto al monitoraggio, se presenta valori di T_i^2 e Q_i superiori ai limiti stabiliti è definito in Fuori Norma.

4.4 Validazione delle Carte di Controllo

La Validazione delle Carte di Controllo consiste nel proiettare i dataset *“Test”* sul modello costruito dal dataset *“Training”* e valutare la capacità delle Carte di individuare situazioni anomale. La validazione consta di due fasi:

1. Validazione NOC (NOC = Lotti in condizioni operative normali)
2. Validazione AOC (AOC= Lotti in condizioni operative anormali)

4.4.1 Validazione NOC delle Carte di Controllo

La validazione NOC del modello è effettuata proiettando dei lotti *“Test”* in condizioni operative normali sul modello costruito in base al dataset *“Training”*, per valutare la capacità di diagnosi del modello di lotti NOC ed individuare il numero di falsi allarmi associati alle carte. Come descritto nel paragrafo 4.3, a partire dai dati reali del dataset *“Test NOC”* sono calcolati i rispettivi *“Score”* sul modello MPCA, i rispettivi valori dei test statistici Q e T^2 che sono confrontati con i valori dei limiti controllo T_{limite}^2 e Q_{limite} ; se il lotto i -esimo sottoposto al monitoraggio presenta valori di T_i^2 e Q_i superiori ai limiti stabiliti è definito in Fuori Norma. Le illustrazioni 28 e 29 riportano le Carte di Controllo T^2 e Q per il dataset *“Test di lotti NOC”*.

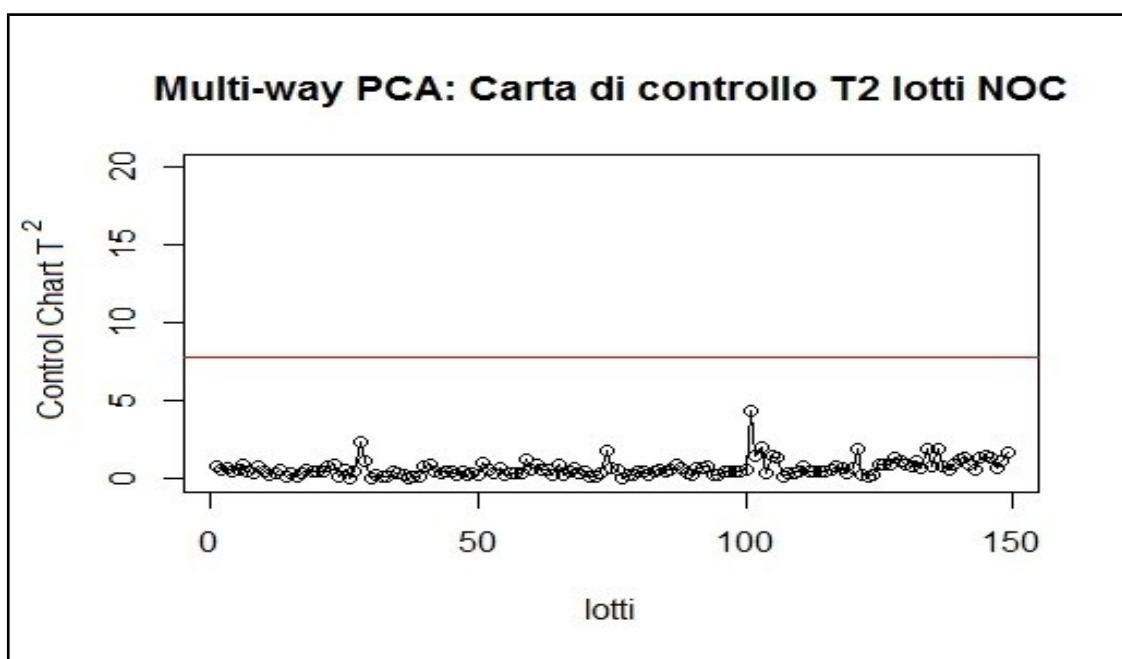


Illustrazione 28: Risultati del Test di Validazione delle Carte: Carta di controllo di Hotelling T^2 del Set di lotti Test NOC (lotti in condizioni operative normali)

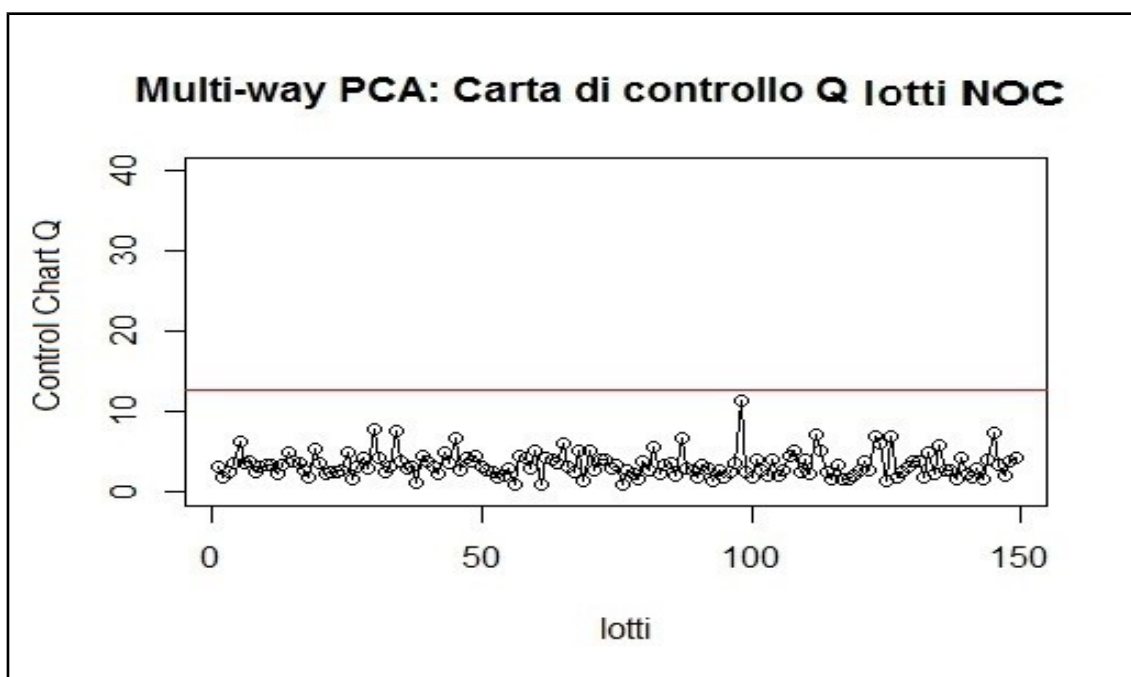


Illustrazione 29: Risultati del Test di Validazione delle Carte: Carta di controllo Q del Set di lotti Test NOC (lotti in condizioni operative normali)

La formula seguente indica la Sensitività NOC, ovvero il rapporto tra lotti “Test NOC” al di sotto del limite di controllo e il numero totale di lotti NOC testati.

$$SENSITIVITA\ NOC = \text{lotti test NOC sotto limite} / \text{test NOC totali}$$

I risultati della Validazione NOC delle Carte di controllo, costruite in base al dataset “Training”, ne evidenziano le ottime prestazioni: la totalità dei lotti NOC utilizzati per il test, sono al di sotto del limite di controllo, per cui sono correttamente classificati dalle Carte come lotti in condizioni operative normali e la Sensitività NOC della Carta è pari al 100%.

	Sensitività NOC del Modello
Per la carta di controllo T2	100,00%
Per la carta di controllo Q	100,00%

Tabella 2: Sensitività: Risultati del Test NOC (Lotti in Condizioni Operative Normali) per verificare le performance delle Carte di Controllo costruite in base al Modello MPCA

4.4.2 Validazione AOC delle Carte di Controllo

La validazione AOC del modello è effettuata proiettando dei lotti “Test AOC” in condizioni operative anormali sul modello costruito sul dataset “Training” per valutare la capacità di diagnosi delle Carte di anomalie associate al lotto ed individuare il numero di mancati allarmi associati al modello. Dai dati originari del Set di lotti “Test AOC” sono calcolati i rispettivi “Score” sul modello MPCA e i rispettivi valori dei test statistici Q e T^2 . Se il lotto i -esimo sottoposto al monitoraggio presenta valori di T_i^2 e Q_i superiori ai limiti di controllo delle Carte è classificato Fuori Norma (AOC). Le illustrazioni 30 e 31 riportano le Carte di Controllo T^2 e Q per il Set dei 150 lotti “Test AOC”.

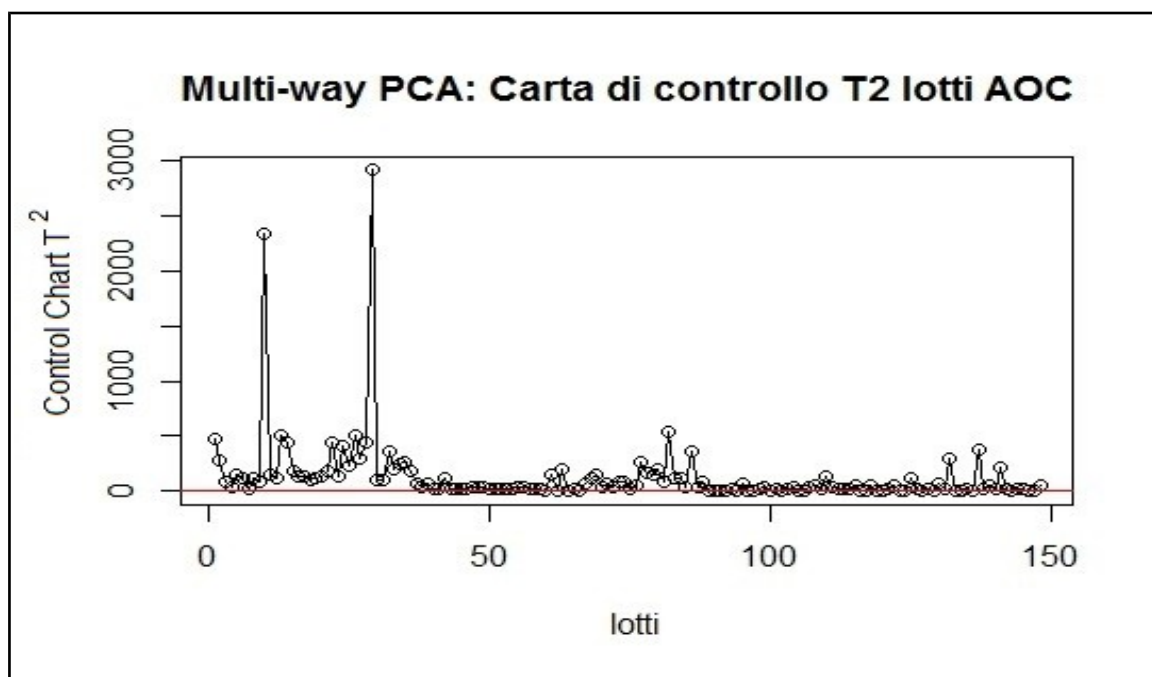


Illustrazione 30: Risultati del Test di Validazione delle Carte: Carta di controllo di Hotelling del T^2 Set di lotti Test AOC (lotti in condizioni operative anormali)

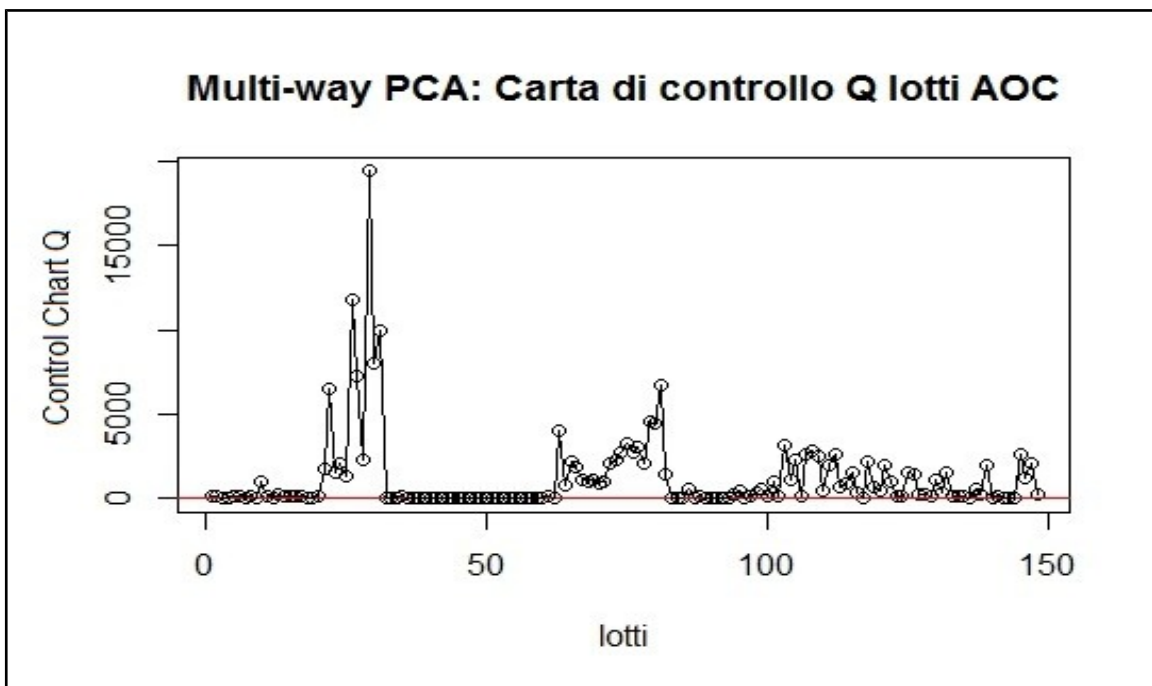


Illustrazione 31: Risultati del Test di Validazione delle Carte: Carta di controllo Q del Set di lotti Test AOC (lotti in condizioni operative anormali)

Il rapporto tra i lotti “*Test AOC*” che superano i limiti e il totale di lotti test AOC è la “*Sensitività AOC*” del modello, parametro che indica la capacità delle Carte di diagnosi di lotti in condizioni operative anormali.

$$SENSITIVITA\ AOC = \text{lotti test AOC sopra limite} / \text{test AOC totali}$$

Il numero di lotti presenti nel dataset “*Test AOC*” è pari a 150, le prestazioni delle Carte, nella diagnosi lotti anormali sono inferiori rispetto a quella di lotti normali. Le carte di controllo presentano un errore pari al 8%, ovvero alle carte di controllo sono associate 12 mancati allarmi (corrispondenti a lotti in condizioni anomale) su 150 lotti.

	Sensitività AOC del Modello
Per la carta di controllo T2	92,00%
Per la carta di controllo Q	94,00%

Tabella 3: Sensitività AOC: Risultati del Test AOC (Lotti Condizioni Operative Anormali) per verificare le performance delle Carte di Controllo costruite in base al Modello MPCA

4.5 Carta di Controllo Riassuntiva

La carta di controllo riassuntiva dà un immediata lettura della statistica T^2 e della statistica Q, dà indicazione della regione di forma rettangolare chiamata

Normal Operation Region (NOR), entro la quale devono ricadere i campioni a norma (Illustrazione 31). Per ogni i -esimo campione T_i^2, Q_i sono i valori delle due statistiche di controllo calcolate. I valori di T_{limite}^2 e Q_{limite} sono i limiti delle due carte di controllo (linee tratteggiate in rosso nell'Illustrazione 31).

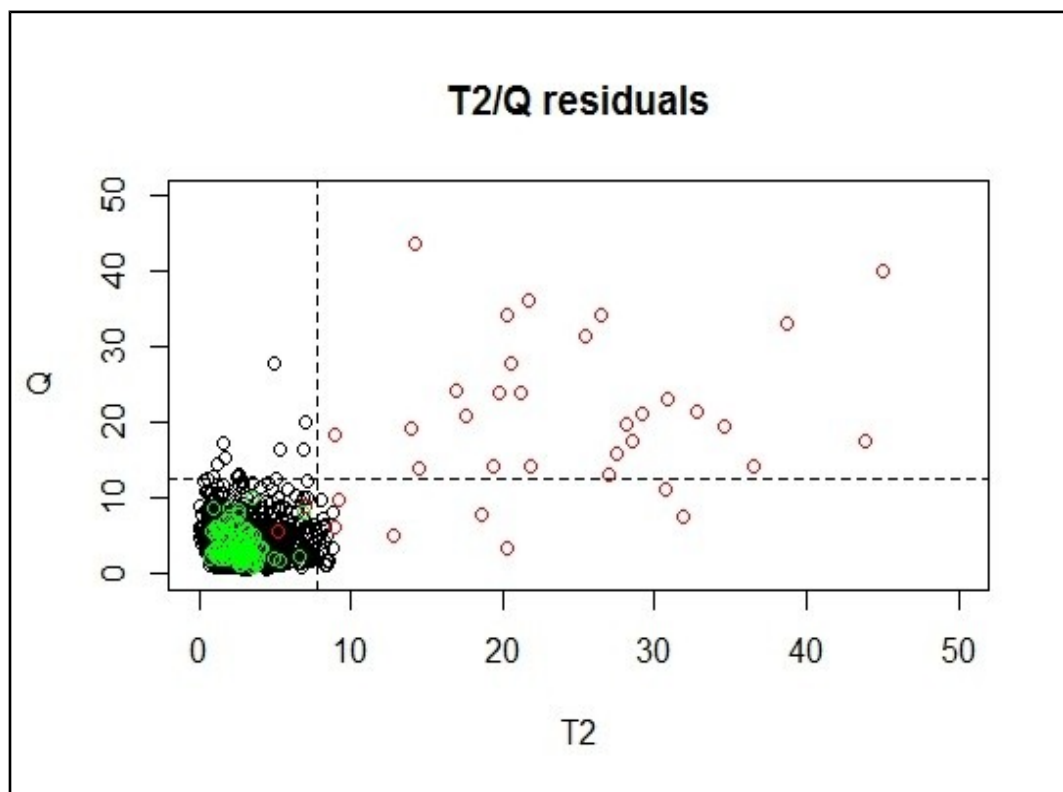


Illustrazione 32: Risultati del Test di Validazione delle Carte: Carte di Controllo Riassuntiva del Set di lotti "Training"(cerchi neri), del Set di lotti Test NOC (lotti in condizioni operative normali,cerchi verdi), e del Set di lotti Test AOC (lotti in condizioni operative anormali,cerchi rossi)

Il grafico presente nell'Illustrazione 31 riporta la carta di controllo riassuntiva dove sono rappresentati i risultati dei test statistici T^2 e Q , per:

- il Set di lotti "Training" sulla quale è stato costruito il modello e le Carte di

Controllo (rappresentato in figura da cerchi neri);

-il Set di lotti "*Test NOC*" costituito da lotti NOC (lotti in condizioni operative normali) utilizzato per testare la capacità di diagnosi le Carte (rappresentato in figura da cerchi verdi);

-il Set di lotti "*Test AOC*" costituito da lotti AOC (lotti in condizioni operative anormali) utilizzato per testare la capacità di diagnosi delle Carte (rappresentato in figura da cerchi rossi);

Ogni lotto presente nell'Illustrazione 31, ha due coordinate (T^2, Q), l'asse X rappresenta i valori della carta di controllo T^2 e l'asse Y rappresenta i valori della carta di controllo Q. Le due linee tratteggiate sono i limiti delle carte di controllo T^2_{limite} e Q_{limite} calcolati in riferimento al modello.

Dalla carta riassuntiva precedente si possono evidenziare le seguenti considerazioni :

- i lotti utilizzati per costruire il modello (cerchi neri) sono posizionati nella regione NOR (Normal Operation Region), in quanto descrivono la variabilità comune presente nei lotti in condizioni operative normali.
- i lotti AOC (cerchi rossi) utilizzati per validare il modello, sono posizionati per la maggior parte nella regione caratterizzata da valori dei test statistici T^2, Q molto elevati, quindi fuori norma, le carte di controllo hanno dimostrato la validità del modello nel riconoscere lotti in condizioni operative anormali.
- i lotti NOC (cerchi verdi) utilizzati per validare il modello, sono posizionati nella stessa regione dove si trovano i lotti su cui è stato costruito il modello, ovvero la regione NOR (normal operation condition).

In conclusione i risultati dei Test effettuati, evidenziano le ottime performance delle Carte di Controllo, nel diagnosticare reali situazioni anomale riferite ai lotti, rappresentative di una variabilità straordinaria, estranea al modello MPCA.

4.6 Individuazione delle Anomalie di Processo e Identificazione delle determinanti di lotti anomali

Quando si realizzano delle situazioni anomale di funzionamento dell'impianto, per cui il processo non è più in condizioni operative normali, si parla di accadimento di un'anomalia. Grazie alla costruzione delle carte di controllo e dei rispettivi limiti, è possibile il monitoraggio di processo, con prestazioni che dipendono dal modello che è stato adottato. La carta di controllo multivariata è uno strumento che ci permette di studiare la natura di variabilità del processo produttivo. L'obiettivo della carta di controllo è quella di monitorare l'output di processo. Il valore anomalo di un parametro dipende da cause straordinarie. Le carte di controllo multivariate qui studiate, sono di tipo unilaterale in quanto presentano un unico limite, il lotto che supera questo limite viene classificato come Fuori Norma. Nei modelli statistici multivariati, si assume che i campioni su cui si costruisce il modello abbiano una distribuzione normale. Dopo aver creato il modello è necessario calcolare i limiti di confidenza del sistema cioè i limiti che il modello ha nelle rappresentazioni dei nuovi campioni. Se il lotto considerato presenta una grossa fluttuazione dei dati e i valori delle carte di controllo T^2 e Q del lotto i -esimo superano i limiti definiti dal modello, il lotto è etichettato come AOC ovvero un lotto operante in condizioni anomale. Per tale lotto è possibile calcolare il contributo al T^2 per determinare quale variabile e in quale istante di tempo è maggiormente responsabile della variazione del nuovo lotto rispetto al modello. Tale contributo viene calcolato con la formula seguente (Kourti, 2005):

$$T_{con}^2 = t_i \Lambda^{-1} p_i^T$$

dove t_i sono gli "Score" del nuovo lotto di dimensioni $1 \times A$ proiettato sul piano

definito dalla A componenti principali del modello. In questa Tesi è stato scelto di selezionare 3 componenti principali, λ sono gli autovalori associati a ciascuna componente principale, p_i^T è la trasposta dei loadings del modello MPCA costruito dal dataset Training. Il grafico dei contributi di ciascuna variabile di processo monitorata, è chiamato *"Contribution Plot"* (Hu & Yuan, 2008). Le seguenti figure riportano le carte di controllo Q e T^2 di una Simulazione di Monitoraggio su dati in riferimento a 120 lotti monitorati in ordine successivo temporalmente, forniti dall'Azienda Mixer.

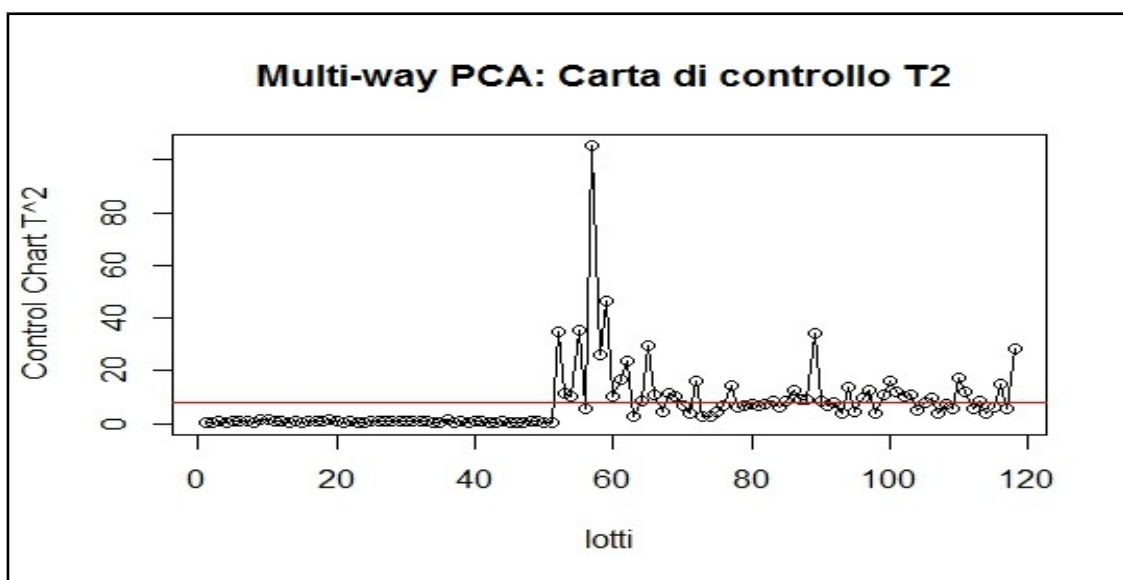


Illustrazione 33: Risultati della simulazione di monitoraggio su dati di processo

forniti dall'azienda Mixer Spa: Carta di Controllo T^2 di 120 lotti

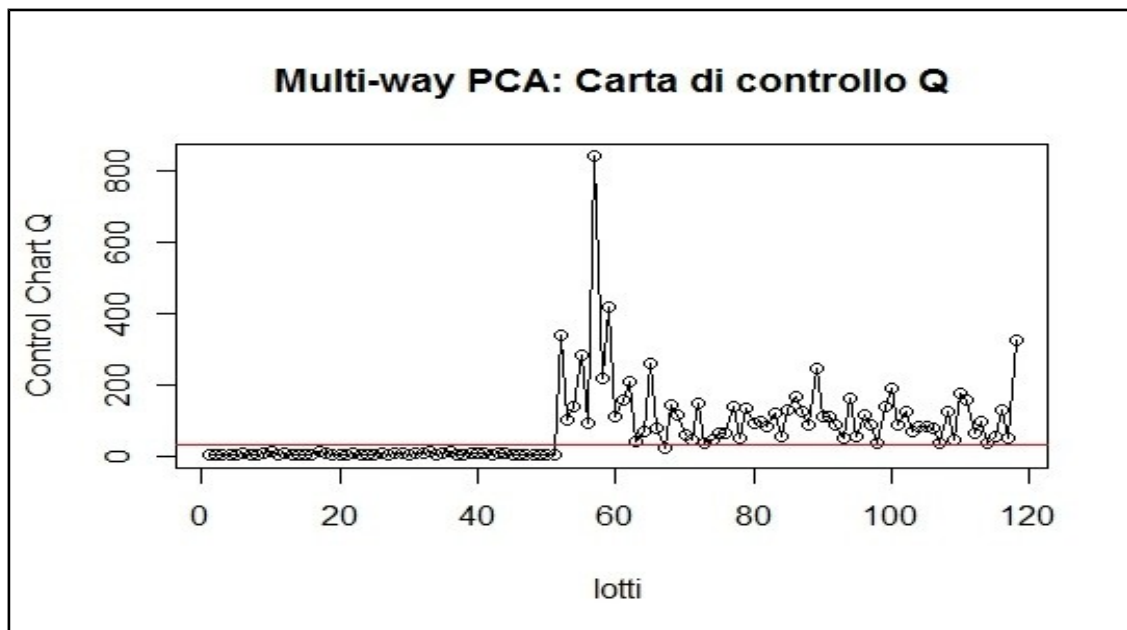


Illustrazione 34: Risultati simulazione monitoraggio sui dati forniti dall'azienda

Mixer Spa: Carta di Controllo di Hotelling T^2 di 120 lotti

La variabilità descritta dal modello è rispettata dai primi 52 lotti, rimasti al di sotto del limite della Carta di Controllo, al livello del 53-esimo lotto è evidente uno scostamento che porta il processo ad essere fuori controllo.

4.6.1 Parametri di processo determinanti della Variabilità Straordinaria

Per i primi 3 lotti anomali ovvero il 53-esimo, il 54-esimo, e il 56-esimo (il lotto con massima variabilità) analizzati sono calcolati i contributi al T^2 , per determinare quale variabile di processo e quale istante di tempo sono maggiormente responsabili della variazione anomala del nuovo lotto

monitorato, rispetto al modello. Il processo di miscelazione analizzato è suddiviso in quattro sottofasce, che sono sequenziali nel tempo. I dati sono campionati per 4 istanti successivi, in corrispondenza di ogni sottofase. Per ogni sottofase sono monitorate 4 variabili di processo che sono: la temperatura della Mescola, la Durata della sotto-fase, la Potenza necessaria per la miscelazione e il Peso della mescola.

Le illustrazione 35,36,37 mostrano i 3 grafici a barre dei Contributi, costituiti da 16 colonne identificative di:

1. **Temperatura della mescola1** nella prima sottofase di miscelazione
2. **Durata1**: intervallo di tempo della prima sottofase di miscelazione
3. **Potenza1**: potenza necessaria nella prima sottofase di miscelazione
4. **Peso1** della mescola nella prima sottofase di miscelazione
5. **Temperatura2** della mescola nella seconda sottofase di miscelazione
6. **Durata2**: intervallo di tempo della seconda sottofase di miscelazione
7. **Potenza2** potenza necessaria nella seconda sottofase di miscelazione
8. **Peso della mescola2** nella seconda sottofase di miscelazione
9. **Temperatura3** della mescola nella terza sottofase di miscelazione
10. **Durata3**: intervallo di tempo della terza sottofase di miscelazione
11. **Potenza3** potenza necessaria nella terza sottofase di miscelazione
12. **Peso3 della mescola** nella terza sottofase di miscelazione
13. **Temperatura4 della mescola** nella quarta sottofase di miscelazione
14. **Durata4**: intervallo di tempo nella quarta sottofase di miscelazione
15. **Potenza4** potenza necessaria nella quarta sottofase di miscelazione
16. **Peso4 della mescola** nella quarta sottofase di miscelazione

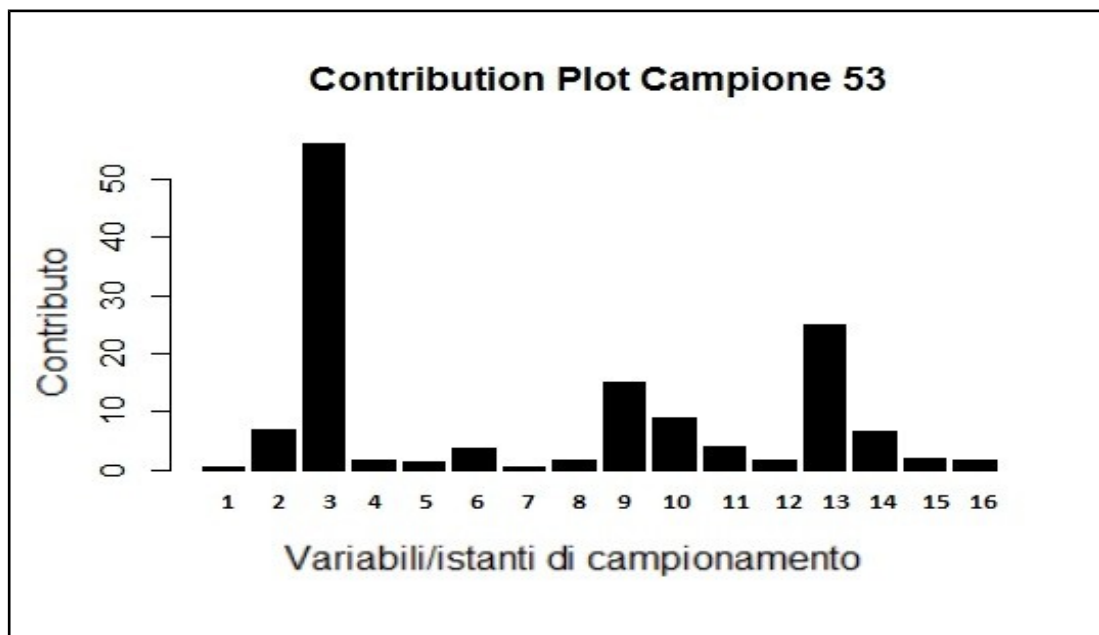


Illustrazione 35: Contribution Plot del Lotto 53-esimo classificato anomalo

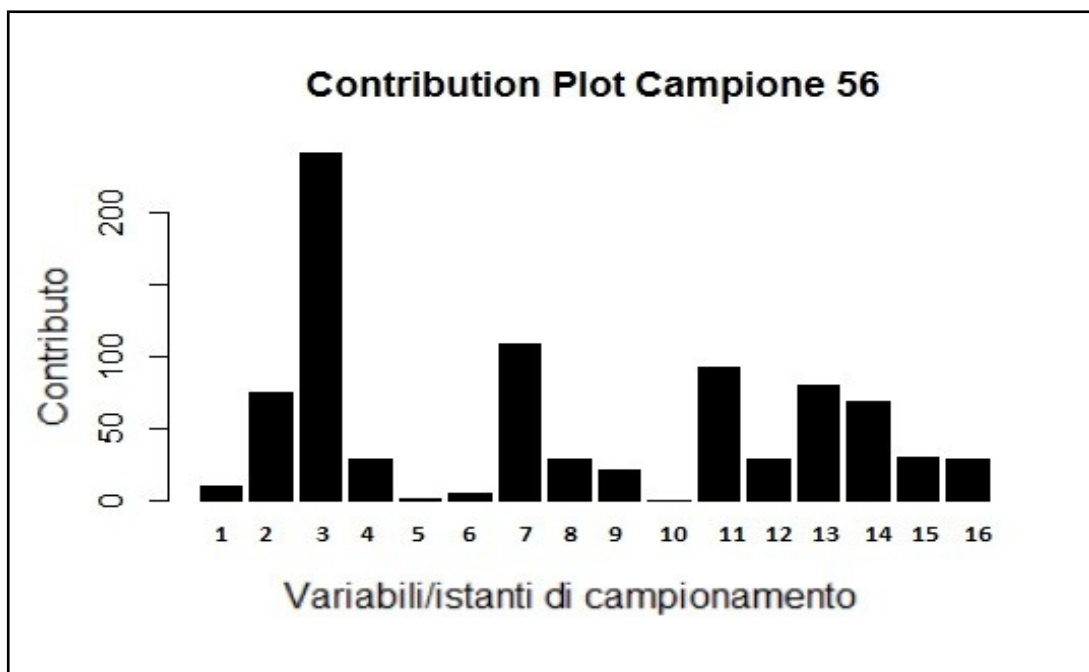


Illustrazione 36: Contribution Plot del Lotto 56-esimo classificato anomalo

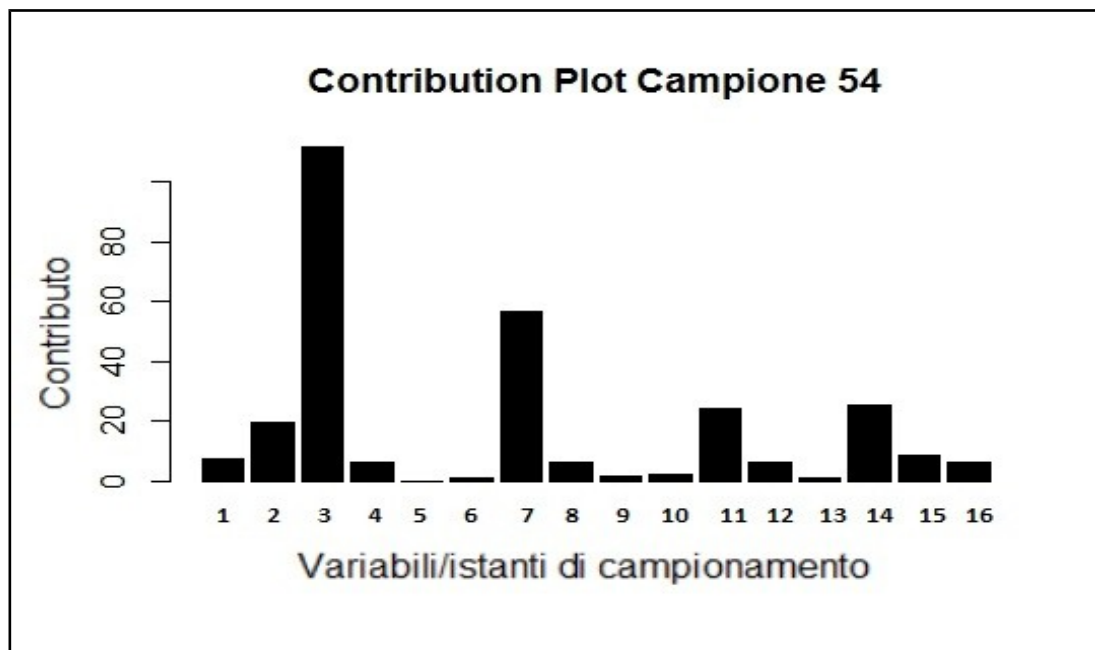


Illustrazione 37: Contribution Plot del Lotto 54-esimo classificato anomalo

Per i 3 lotti analizzati, i contributi al T^2 evidenziano che la variabile di processo della Potenza monitorata al termine della prima sottofase di Miscelazione, sia responsabile della variazione straordinaria dei suddetti lotti rispetto al modello. In altre termini la grossa fluttuazione visibile nelle carte di controllo T^2 a partire dal 53-esimo lotto è imputabile (illustrazione 32,33) ad una variazione della Potenza necessaria per la miscelazione, rispetto la variabilità comune al modello. Nell'esempio precedente i lotti sono classificati come fuori norma se superano i limiti delle carte di controllo, un ulteriore modo per evidenziare una situazione anomala è individuare eventuali andamenti sistematici dei lotti (trend), che sono monitorati uno dopo l'altro nel tempo. Altre situazioni anomale si presentano nel caso in cui si evidenzi un andamento ciclico dei dati dovuto a delle cause ricorrenti (un ambiente climaticamente non controllato) o nel caso in cui i lotti monitorati realtime tendano ad avvicinarsi progressivamente al limite di controllo (fenomeno detto Deriva illustrazione 38)

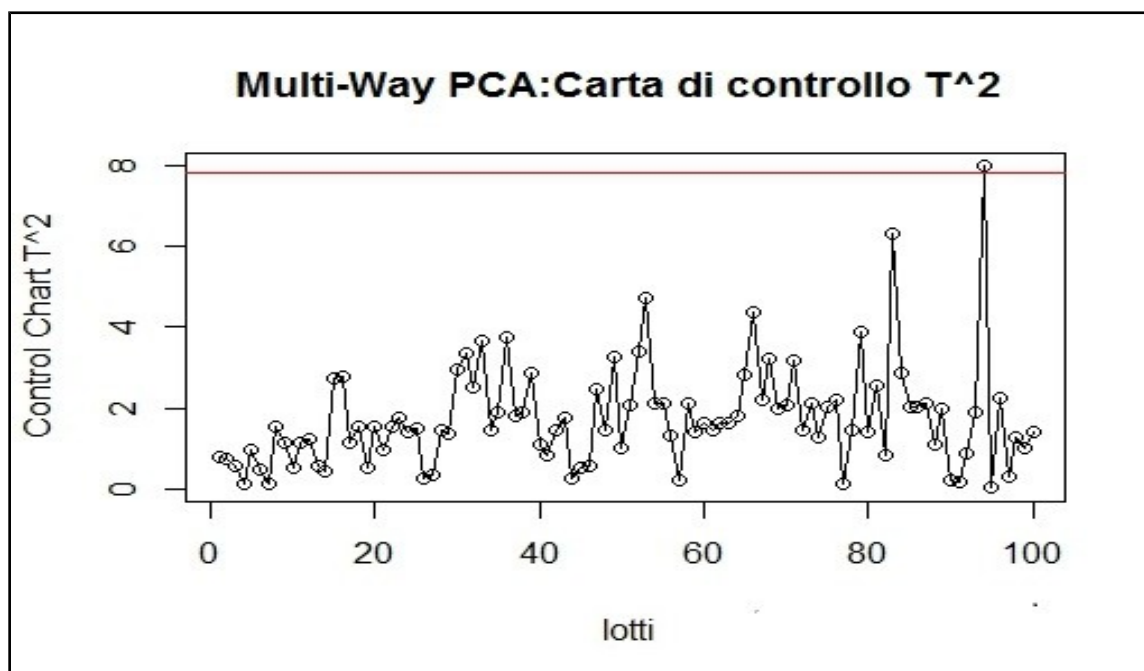


Illustrazione 38: Carta di controllo T^2 di lotti monitorati in ordine successivo nel tempo con un andamento ciclico indicativo del fuori controllo del processo (fenomeno chiamato Deriva)

La carta di controllo nell'illustrazione 37, mostra come i lotti monitorati successivamente non seguano un andamento costante, ma sia presente un effetto “stagionale” che porta il processo a superare il limite di controllo.

La carta di controllo riconosce un lotto fuori norma, nell'istante in cui il lotto supera i limiti, spetta all'analista che si occupa di interpretare le carte di controllo al termine della produzione o all'operatore che si occupa del monitoraggio online, di riconoscere dei trend positivi, che possono portare il processo fuori controllo e di conseguenza intervenire analizzandone le determinanti della variabilità straordinaria, eventualmente ritirando il lotto se danneggiato e verificando che le condizioni di processo rispecchino le condizioni ottimali altrimenti intervenire per ripristinarle. Le suddette fasi sono descritte nel capitolo 5.

4.6.2 Individuazione delle correlazioni tra le variabili

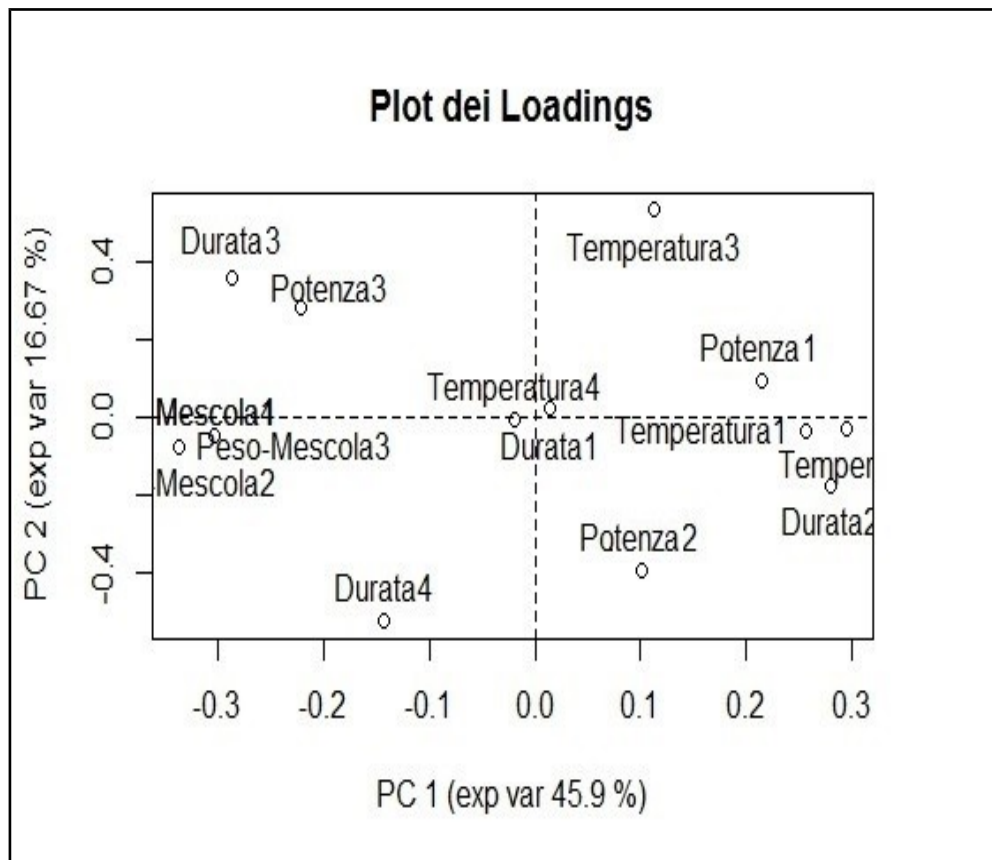


Illustrazione 39: Proiezione delle variabili di processo originarie nello spazio dimensionale definito dalle prime due componenti principali del modello MPCA costruito dal dataset di lotti Training

Nell'Illustrazione 39 (Plot dei Loading) è possibile individuare la proiezione delle variabili originali di processo ovvero la Temperatura, la Potenza, la Durata e il Peso della miscela monitorate nei 4 istanti di campionamento, sul piano definito dalle componenti principali del modello Multi-way PCA. Il piano delle due componenti principali permette di rappresentare la nuvola di punti,

rappresentativa dei lotti, nella prospettiva che massimizza la varianza dei dati. In questo caso proiettando le vecchie variabili è possibile notare le correlazioni esistenti tra i parametri di processo. Ad esempio le informazioni interpretabili dal grafico seguente (Illustrazione 39) sono che il peso mescola è costante nei 4 istanti di campionamento e che la durata della terza sottofase sia strettamente correlata alla potenza necessaria per la miscelazione.

Capitolo 5

Monitoraggio online e Aggiornamento del modello

I modelli statistici multivariati possono essere utilizzati per il monitoraggio di processo in linea allo scopo di identificare anomalie, che se non rilevate, pregiudicano la qualità finale del prodotto. I dati storici del processo di miscelazione trattato, sono stati usati con approccio di tipo Multiway PCA, in cui il modello è tarato sullo storico del processo e poi utilizzato per il monitoraggio del sistema produttivo e per l'individuazione di eventuali anomalie. Il modello Multiway in prima istanza, può essere utilizzato per il monitoraggio online, i risultati possono essere interpretati, post-produzione, anche per classificare le anomalie di processo e restringere il campo di ricerca delle cause scatenanti. Il metodo proposto ha l'obiettivo di estendere il controllo durante l'esecuzione del processo di miscelazione in aggiunta al controllo di qualità post-produzione già presente nell'Azienda Mixer. Il modello Multiway PCA può essere costantemente aggiornato con l'inserimento di dati di nuovi lotti per essere nel tempo aggiornato e robusto.

5.1 Simulazione degli Step di Controllo Statistico Online

Il modello MPCA è stato costruito e validato allo scopo di essere utilizzato per il monitoraggio online tra un batch e l'altro. Al termine della fase produttiva di miscelazione che consta di quattro istanti temporali, un lotto può essere classificato come NOC (lotto in condizioni operative normali) se rimane sotto i limiti di controllo altrimenti come AOC (lotto in condizioni operative anormali). Se tale lotto è classificato come NOC procede nelle fasi successive di lavorazione, se invece il lotto è classificato come AOC, la prima azione da effettuare è analizzare le determinanti della variabilità straordinaria come descritto nel capitolo 4 ed eventualmente ritirare il lotto dalla produzione. Se è presente un gruppo di lotti che si avvicinano al limite con un andamento sistematico e crescente è necessario controllare i parametri di processo e attuare delle azioni correttive.

5.1.2 I dati dei Lotti in input per Monitoraggio Online

I dati che devono essere acquisiti in input per monitorare se un lotto è fuori norma, sono le misurazione di quattro parametri di processo (temperatura della mescola, durata della sottofase, potenza elettrica, peso mescola) nei quattro istanti di campionamento successivi. E' stato scelto di effettuare un controllo al termine dell'ultimo istante campionato per ogni lotto, perché la durata totale del processo di miscelazione è molto breve (nell'ordine di secondi). Perciò non è possibile conoscere in ogni istante di campionamento se il lotto i -esimo è fuori limite ma tale informazione è ottenibile al termine dell'ultima sottofase di miscelazione. E' possibile conoscere quale variabile è responsabile dell'anomalia

tramite l'utilizzo dello strumento del “*Contribution Plot*”. I dati del nuovo lotto in input alla carta di controllo, sono disposti in una matrice tridimensionale, dove I è il numero dei lotti, J il numero delle variabili e K gli istanti di campionamento. Ciò significa che ogni sezione orizzontale è una matrice che rappresenta i profili temporali per un singolo lotto, mentre le sezioni verticali rappresentano i valori misurati di ciascuna variabile in un istante temporale K. Per tener conto dell'andamento temporale delle variabili è possibile costruire una matrice bidimensionale mediante un “*Unfolding*” (ovvero apertura della matrice in modo da trasformarla da tridimensionale a bidimensionale) nella direzione dei lotti. Perciò i dati in input di un nuovo lotto si presentano come una stringa di dimensioni 1×16 ($16 = 4 \text{ variabili} \times 4 \text{ istanti di campionamento}$), ottenuta dall'“*Unfolding*” (metodo descritto nel paragrafo 3.7) della matrice 4×4 VariabilixTempo del nuovo lotto i-esimo, nella direzione del lotto. La stringa dei dati riferiti ad un lotto i-esimo è riportata nella tabella 5.1:

Lotti	Temperatura1	Durata1	Potenza1	Peso1	Temperatura2	Durata2	Potenza2	Peso 2	Temperatura3	Durata3	Potenza3	Peso 3	Temperatura4	Durata4	Potenza4	Peso 4
1	962	2	9	173.8	891	42	93	173.8	1091	22	145	173.8	1409	34	198	173.8

Tabella 4: Matrice dei dati riferita al nuovo lotto i-esimo, in input alle carte, per il monitoraggio online: dimensioni 1×16 (4 variabilix4 istanti)

5.1.3 Step della Simulazione di Controllo Statistico Online tra un lotto e il successivo

Il monitoraggio di processo e la diagnosi dei guasti sono compiti importanti che determinano il successo del processo e la qualità del prodotto finale. Nel processo batch, piccoli cambiamenti nelle condizioni operative possono influire

sulla qualità del prodotto finale, che di solito viene esaminata unicamente offline in un laboratorio. Pertanto, un precoce rilevamento dei guasti online è vantaggioso per prendere misure correttive prima che il processo sia compiuto e quindi impedire guasti per i successivi lotti (Shen, Ge, & Song, 2015).

L'illustrazione 39 mostra gli Step del controllo online di processo.

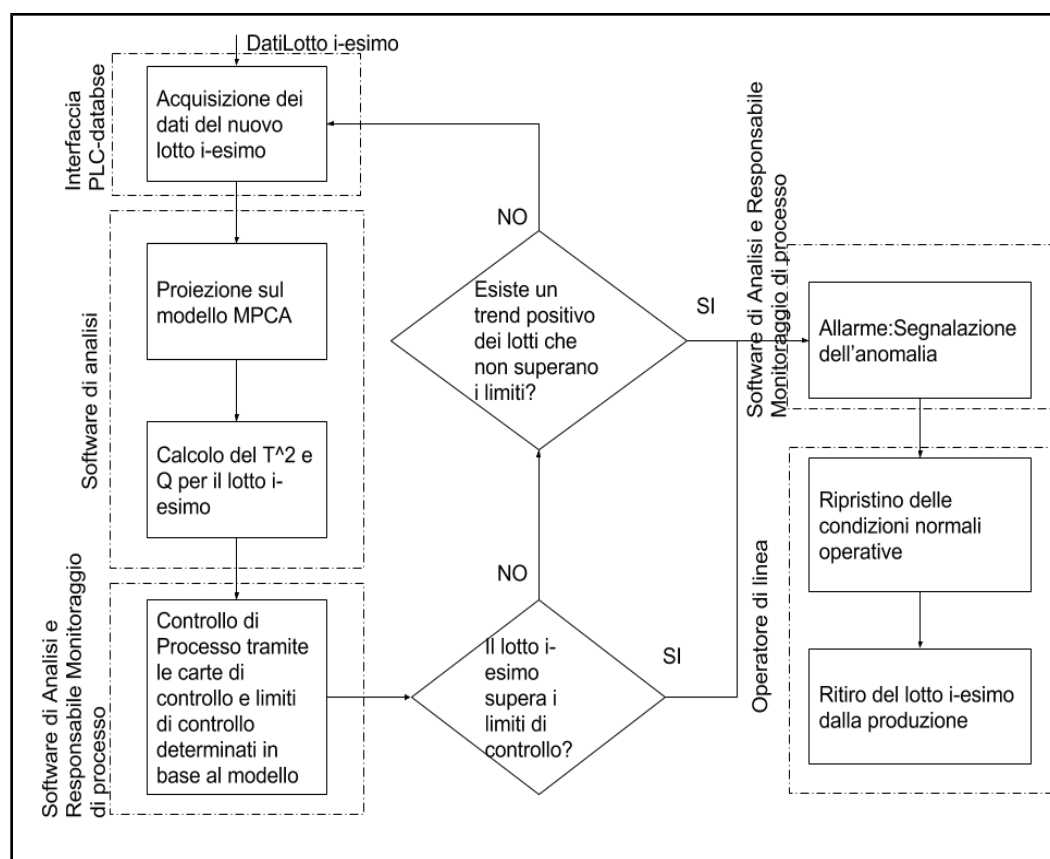


Illustrazione 40: Step del Controllo Online di Processo

Gli Step della Simulazione di Controllo Statistico Online sono riportati successivamente:

1. Acquisizione dei dati del nuovo lotto i-esimo al termine degli istanti di campionamento del processo.
2. Proiezioni nel modello MPCA (precedentemente calcolato sul dataset).

“Training”)

3. Calcolo del T_i^2, Q_i per il lotto i-esimo
4. Controllo statistico mediante carta di controllo T^2, Q e i limiti di controllo determinata sul dataset “Training”.
5. Classificazione del nuovo lotto come lotto NOC (in condizioni operative normali) o come lotto AOC (in condizioni operative anormali). Se il lotto è NOC, si valuta la presenza di un trend positivo di lotti che si avvicinano al limite, se non è presente nessun trend si ritorna allo Step 1 e si procede con il monitoraggio del lotto successivo. Se il lotto è AOC o se è presente un trend positivo di lotti che si avvicinano al limite di controllo, si segnala un'anomalia tramite un Allarme: si analizzano le determinanti della variabilità e controllano i parametri di processo ed infine si ritira il lotto dalla produzione affinché possa essere successivamente rilavorato. Eseguite questi Sotto-Step, si ritorna allo Step 1 e si monitora il lotto successivo.

Il controllo delle Carte di Controllo e l'analisi dei parametri di processo vengono effettuati da un addetto Responsabile di Monitoraggio che si occupa della supervisione del processo di miscelazione tramite un dispositivo hardware localizzato nel laboratorio di Controllo di Qualità che si trova all'interno dello stabilimento. Il modello segnala l'allarme nel momento in cui il lotto supera i limiti, ma il Responsabile di Monitoraggio di Processo è tenuto a segnalare una situazione anomala anche quando riconosce un trend positivo dei lotti che si avvicinano al limite di controllo (fenomeno chiamato “Deriva”). Nel momento in cui è segnalata un'anomalia, questa informazione è trasmessa real-time dal laboratorio alla linea produttiva tramite dei dispositivi di allarme; l'operatore di linea dovrà ritirare il lotto i-esimo segnalato come anomalo e registrarlo nell'apposito database di lotti AOC (lotti in condizioni operative anomale).

5.2 Impatto organizzativo e Risorse aziendali necessarie per l'introduzione del Sistema di Monitoraggio Online per il processo di Miscelazione

Il Controllo introdotto al Processo di Miscelazione è cosiddetto in *"process"*, ovvero il monitoraggio avviene mentre l'attività da controllare è in corso di svolgimento, tramite l'utilizzo di un software, creato ad hoc, che genera le carte di controllo multivariate, costruite sul modello MPCA. Il software di controllo, basato sul modello MPCA, costruito sul dataset *"Training"*, come descritto nei capitoli precedenti, è stato testato prima della sua applicazione tramite due dataset per conoscere la capacità di diagnosi di anomalie delle carte di controllo e il numero di falsi allarmi associati al sistema di controllo.

5.2.1 Impatto Organizzativo per lo sviluppo e l'attuazione del Sistema di Monitoraggio

L'introduzione di un nuovo sistema di monitoraggio ha un impatto organizzativo ad ampio raggio, coinvolgendo i processi direzionali, operativi, di controllo e la gestione delle risorse dell'Azienda. Nell'Illustrazione 40 è proposto uno schema che mostra la pianificazione dell'introduzione del sistema di monitoraggio online del processo di miscelazione.

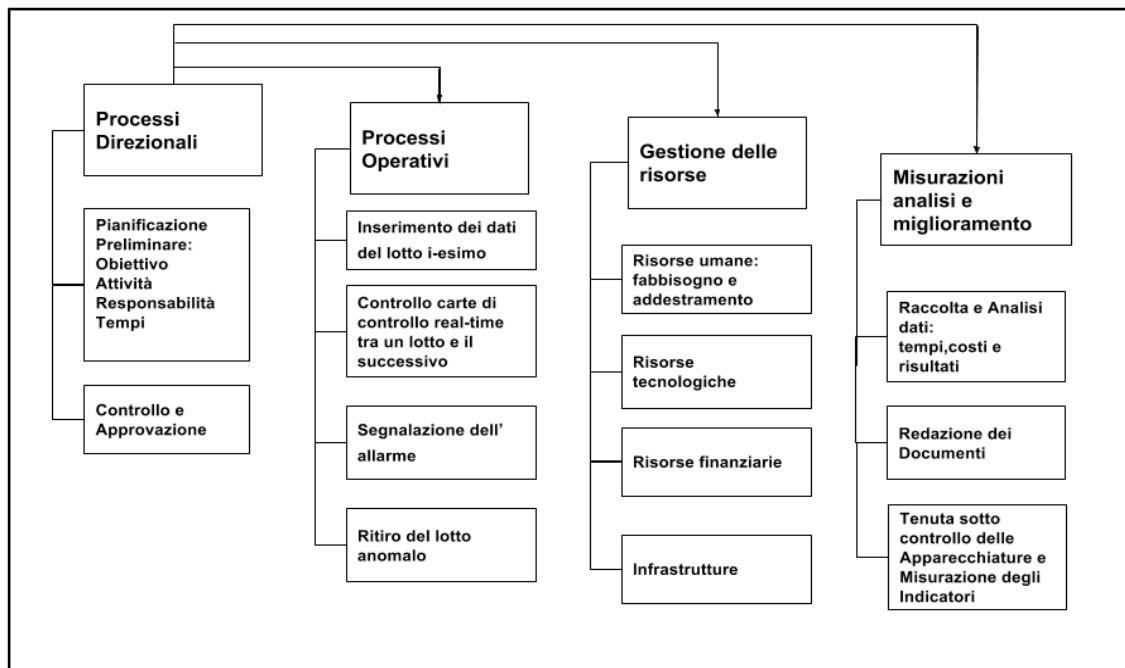


Illustrazione 41: Pianificazione del Sistema di monitoraggio per il Processo di Miscelazione

5.2.2 Processi Direzionali

I processi direzionali per lo sviluppo e l'attuazione del sistema di monitoraggio includono:

- Pianificazione preliminare

L'introduzione di un sistema di monitoraggio per il processo di miscelazione ha come obiettivo, l'analisi dei parametri di processo per differenziare i prodotti tra loro, identificare alterazioni tra i batch, per individuare i parametri responsabili delle peculiarità riscontrate e per aumentare il know-how organizzativo. La direzione stabilisce le attività e i tempi necessari per lo sviluppo e l'attuazione del sistema di monitoraggio e assicura che siano definite le responsabilità e

comunicare all'interno dell'organizzazione .

- Controllo

La direzione attua un controllo sui risultati del monitoraggio rispetto quanto è stato pianificato.

5.2.3 Processi Operativi

I Processi Operativi del sistema di monitoraggio online includono:

- Inserimento all'interno del software di monitoraggio tramite un interfaccia, dei dati di ciascun lotto, che realtime sono misurati dai sensori presenti sul macchinario di Miscelazione.
- Controllo delle carte di controllo realtime costruite dal Software. E' necessario che il Responsabile di Monitoraggio verifichi che il risultato per ciascun lotto rientri all'interno dei limiti di controllo.
- Segnalazione dell'allarme, nel momento in cui il lotto superi i limiti delle carte oppure nel momento in cui l'operatore riscontri il fenomeno di deriva dei lotti
- Ritiro del lotto se è presente un'anomalia: il processo è di tipo batch quindi alla fine dell'ultima sottofase di miscelazione un operatore deve occuparsi del ritiro del lotto che ha scatenato l'allarme. Altrimenti si prosegue con la produzione del lotto successivi.

5.2.4 Gestione delle Risorse

L'introduzione e l'attuazione del Sistema di Monitoraggio necessita la messa a disposizione di infrastrutture, di risorse umane, tecnologiche e finanziarie.

Lo sviluppo del sistema di monitoraggio è stato progettato per creare un'interazione tra diversi database e una serie di interfacce che sono mostrate nell'illustrazione 41.

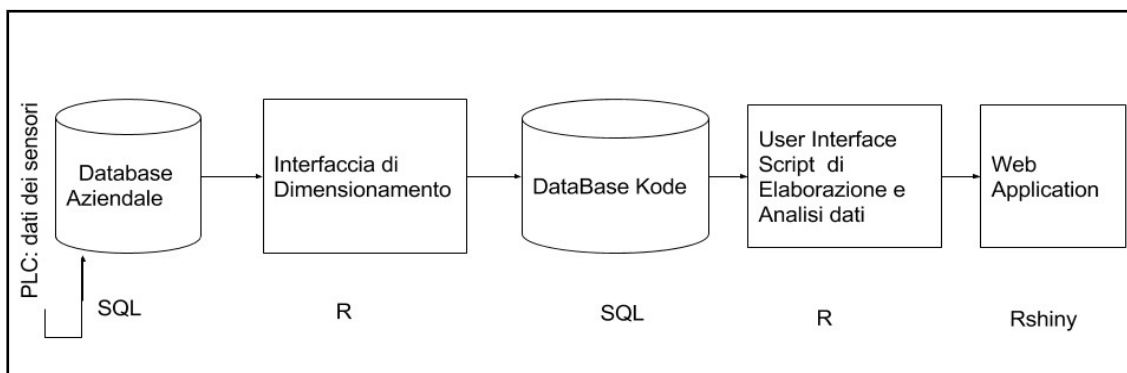


Illustrazione 42: Sviluppo del Sistema di Monitoraggio: Database e Interfacce

I dati di processo provenienti da sensori posti lungo la linea produttiva sono elaborati da un PLC Programmable Logic Control, un controllore per industria specializzato nella gestione dei controlli dei processi industriali. I dati di processo sono conservati all'interno di un database aziendale in linguaggio SQL. Un interfaccia di dimensionamento costituita da una serie di script (programmi scritti in uno specifico linguaggio di programmazione) in linguaggio di programmazione R, riceve in input i dati del database aziendale, li ridimensiona secondo le specifiche pianificate e individua eventuali problemi di gestione dati all'interno del database aziendale. I dati rielaborati sono inviati ad un database gestito dall'azienda Kode, fornitrice del servizio di consulenza scientifica, che li elabora con altri script in linguaggio R, ottenendo le carte di controllo multivariate che sono state oggetto di Tesi. Infine è realizzata una User Interface costituita da una Web Application, creata tramite il package Rshiny di Rstudio, che permette di visualizzare su un dispositivo hardware presente nel laboratorio di qualità dell'azienda Mixer, le carte di controllo online del processo di

miscelazione. Il Responsabile di Monitoraggio di Processo ha il compito di individuare anomalie durante la produzione e inviare un allarme lungo la linea produttiva tramite dei dispositivi di segnalazione all'operatore di linea che ha il compito di intervenire ritirando il lotto anomalo.

L'azienda Mixer deve impegnarsi per lo sviluppo del sistema di monitoraggio, sostenendo un investimento iniziale per la consulenza scientifica apportata dall'azienda consulente Kode e deve avere a disposizione delle infrastrutture necessarie per realizzare una postazione di monitoraggio.

Per quanto riguarda le risorse tecnologiche, l'Azienda Mixer si dovrà dotare di un server aziendale dove saranno implementate le interfacce di dimensionamento, il database di analisi dati e la user interface, permettendo un internalizzazione dell'interno sistema di controllo che nella fase iniziale dell'introduzione, sarà gestita esternamente dall'azienda di consulenza Kode. E' prevista una fase di collaudo con dati online dei processi produttivi, in cui è presente l'assistenza dell'azienda di consulenza. Per quanto riguarda le risorse umane è richiesta la definizione di un Responsabile ICT, che si occupi della manutenzione hardware e software del sistema e la definizione di un amministratore del sistema (responsabile del Monitoraggio di Processo) che deve essere formato per tale ruolo. L'azienda di consulenza ha previsto di effettuare delle sessioni di formazione sull'utilizzo delle interfacce e delle carte di controllo multivariate.

5.2.5 Misurazione, Analisi e Miglioramento del Sistema di Controllo

Le attività predisposte per il monitoraggio sono:

- Raccolta e Analisi dati: è prevista la raccolta dei tempi, dei costi di lavorazione, dei parametri di processo associati ai lotti e analizzati post-produzione per la ricerca delle cause scatenanti di anomalie e per effettuare una valutazione

economica post-produzione.

- Redazione documenti: è prevista la redazione di un report che registri ogni lotto come conforme o meno: i dati relativi ad ogni lotto classificato come conforme sono inseriti nel database dei lotti normali e i dati relativi a lotti classificati come anormali sono inseriti nel database lotti anormali.

- Tenuta sotto controllo del sistema di monitoraggio:

Per assicurare la validità dei risultati, le apparecchiature di controllo sono valutate tramite:

- 1) La taratura ad intervalli regolari delle apparecchiature di misurazione, a fronte di campioni di misura riferibili a campioni internazionali. Lo stato di taratura è identificato in un apposito documento e le apparecchiature di misurazione sono protette dal danneggiamento e deterioramento. Sono mantenute le registrazioni della taratura.

- 2) Validazione del software di controllo e della capacità di soddisfare l'applicazione prevista tramite un test prima della sua utilizzazione. Il test prevede per ogni unità funzionale l'inserimento di dati in input di cui già si conoscono gli output, in questo modo si valutano le discrepanze tra i risultati attesi e i risultati ottenuti.

5.3 Monitoraggio offline: Interpretazione delle carte di controllo

I risultati del monitoraggio online, presenti nelle carte di controllo costruite per il processo di Miscelazione, possono essere reinterpretrati per classificare le anomalie di processo e restringere il campo di ricerca delle cause scatenanti.

5.3.1 Diagnosi delle variabili di processo determinanti di anomalie

Il controllo di qualità multivariato fornisce una maggiore capacità di diagnosi delle cause assegnabili alle anomalie di processo e di prodotto rispetto le carte di controllo univariato. Interrogando il modello Multiway-PCA nel punto in cui un evento anomalo è stato rilevato, è possibile estrarre i diagrammi diagnostici di contribuzione (*"Contribution Plot"*) che rilevano il gruppo di variabili di processo che incidono maggiormente sui valori delle carte di controllo Q e T^2 . A differenza del caso univariato, il modello MPCA descrive la correlazione che è presente tra le variabili di processo, ovvero evidenzia il gruppo di variabili che determinano il lotto non conforme. Una causa ipotetica di anomalia può andare a incidere su diversi parametri di processo contemporaneamente, quindi un'analisi univariata non sarebbe appropriata, non ottimizzante perché va ad analizzare delle informazioni ridondanti e non considera la correlazione tra le variabili di processo. I diagrammi non diagnosticano univocamente la causa che determina la variabilità straordinaria, ma forniscono delle informazioni che risultano essere molto utili per restringere il campo della ricerca delle cause. L'interpretazione dei dati di processo offline ovvero al termine del ciclo produttivo, è molto importante per andare a ricercare le cause delle anomalie di processo che incidono sulla qualità finale del prodotto, al fine di intervenire per migliorare la capacità del processo di produrre lotti conformi e la capacità del modello di monitoraggio, di segnalare reali situazioni anomale diminuendo il numero dei falsi allarmi.

5.3.2 Monitoraggio di Processo offline

Al termine del controllo online è necessario andare ad interpretare i dati dei lotti che sono stati prodotti, per riuscire ad individuare diverse situazioni anomale:

- trend di lotti che si avvicinano ai limiti di controllo;
- singoli lotti outliers (al di sopra dei limiti);
- gruppi di lotti di outliers;

Il trend di lotti e i gruppi di lotti che sono al di fuori del limite individuano un "Anomalia di Processo" e il singolo lotto che supera il limite va a definire quella che è chiamata "Anomalia di Prodotto". Questa informazione non è assolutamente irrilevante ma permette di conoscere se la variabilità straordinaria ha determinato un'anomalia che ha prodotto un singolo lotto non conforme oppure ha compromesso la capacità del processo di creare prodotti conformi. L'interpretazione di queste informazioni non è banale: può restringere il campo di ricerca delle cause e suddividerle in cause che si ripercuotono solo sul prodotto *i*-esimo in lavorazione o in cause che possano rendere inefficace l'intero processo. I possibili scenari che si possono presentare dall'analisi delle Carte di Controllo sono quattro e sono riportati nei grafici seguenti:

1) **Scenario 1:** Il processo è sotto-controllo quando il lotto i -esimo si trova al di sotto del limite di confidenza insieme ai lotti che lo precedono.

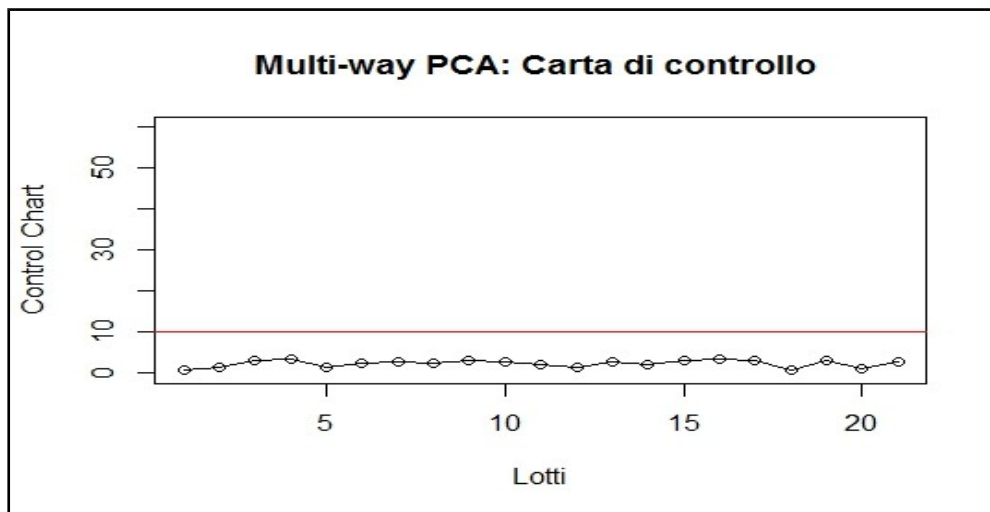


Illustrazione 43: Scenario 1: Carta di Controllo di un set di lotti in normali condizioni operative

2) **Scenario 2:** E' presente un'anomalia di processo se i lotti sono al di sotto del limite ma presentano un trend positivo.

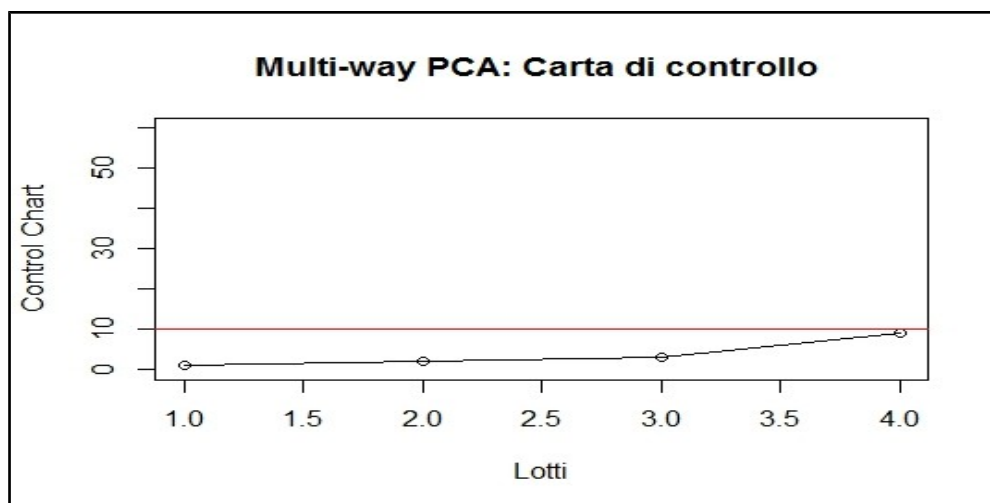


Illustrazione 44: Scenario 2: Carta di Controllo di un set di lotti con trend positivo indicativo di anomalia

3) Scenario 3: è riscontrabile un'anomalia del prodotto se il lotto i -esimo si trova sopra del limite e i lotti precedenti e successivi sono al di sotto del limite.

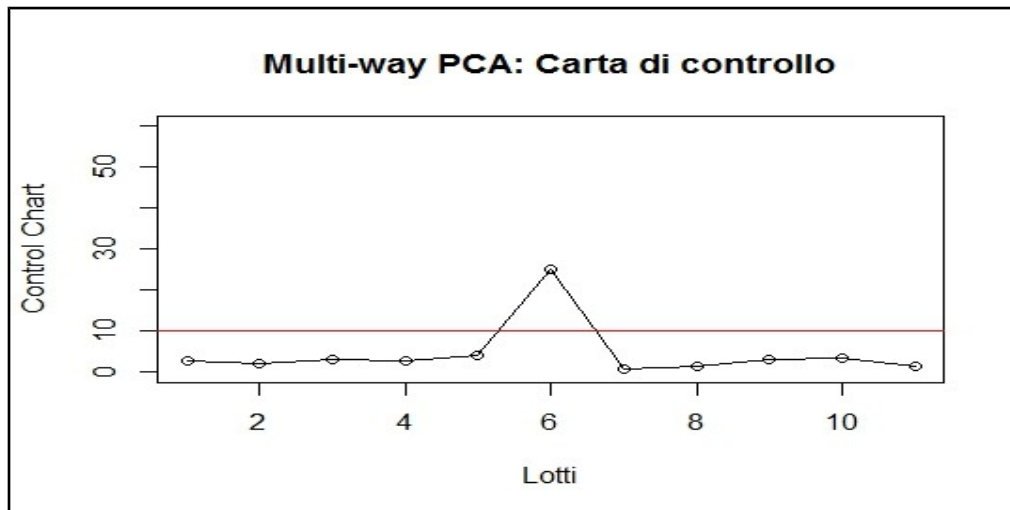


Illustrazione 45: Scenario 3: Carta di controllo di un set di lotti con la presenza di un solo un lotto anomalo

4) Scenario 4 : è riscontrabile un'anomalia del processo se il lotto i -esimo si trova sopra del limite e sia i lotti precedenti e successivi sono sopra del limite.

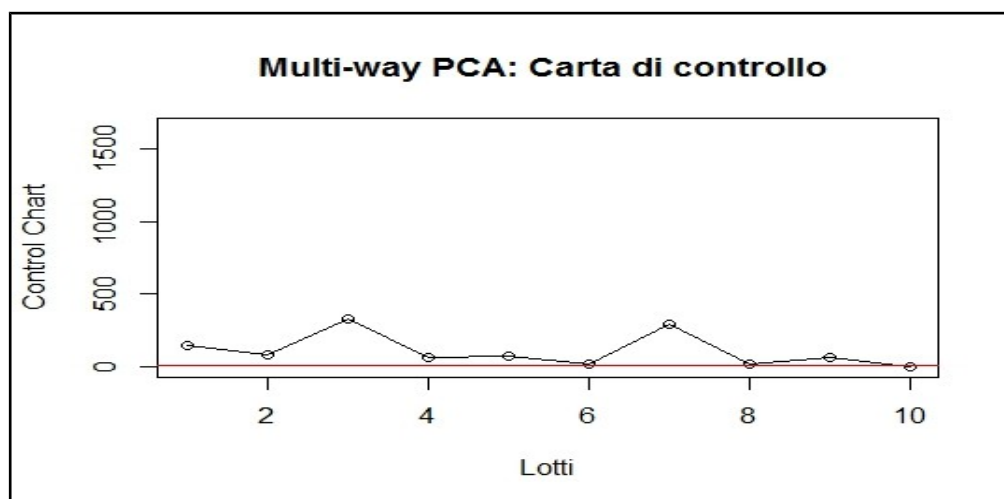


Illustrazione 46: Scenario 4: Carta di controllo rappresentativa di un'anomalia di Processo

Questa interpretazione delle carte di controllo può dare una prima suddivisione delle determinanti che danneggiano il singolo prodotto o portano fuori controllo l'intero processo.

5.4 Aggiornamento del Modello MPCA

L'aggiornamento del modello MPCA consiste nell'inserimento di dati di un nuovo lotto NOC all'interno del dataset *"Training"*. Il dataset *"Training"* è un database che racchiude i dati dei lotti rappresentativi delle condizioni operative normali su cui è costruito il modello MPCA. Questo approccio è di tipo ricorsivo permette di potenziare le capacità di diagnosi delle Carte di Controllo e di diminuire la percentuale di errore del modello.

Il nuovo lotto classificato come conforme (al di sotto dei limiti di controllo) può essere inserito all'interno del Modello; la verifica da fare sul gruppo di lotti conformi da inserire nel modello è il controllo della presenza di un trend crescente tra i lotti, qualora sia presente devono essere scelti altri batch da inserire nel dataset training. L'inserimento di dati relativi a nuovi lotti ha una serie di vantaggi:

1. Il modello viene arricchito di nuovi dati rappresentativi delle variabilità comune del processo, rendendolo più affidabile e rappresentativo della realtà;
2. L'aggiornamento dei dati permette di sostituire ai vecchi dati informazioni più recenti, qualsiasi cambiamento del processo o variazione delle modalità produttive può essere inglobato nel modello;
3. L'aggiornamento del modello non solo può riguardare inserimento di nuovi lotti, ma anche l'inserimento di un nuovo parametro di processo che si vuole monitorare. In questo modo è possibile creare un nuovo modello che valuti la correlazione tra le vecchie variabili e il nuovo parametro di processo

inserito. L'aggiornamento viene effettuato tipicamente offline ma potrebbe essere effettuato anche online se il modello sia già rodato da molto tempo. Il modello MPCA può crescere a pari passo della crescita del processo. L'aggiornamento riguarda anche i limiti di controllo, che come è stato descritto nel capitolo 3 sono calcolati con le formule di Jackson e Mudholkar per la carta di controllo Q e con la formula di Jackson per la carta di controllo T^2 .

5.4.1 Aggiornamento Modello: Inserimento di lotti nel Set Training

Si assuma di avere un modello costruito su j lotti NOC. Per il nuovo campione si esegue la seguente procedura:

1. Pre-processare il nuovo lotto i -esimo, sottraendo e dividendo ogni elemento della stringa dei dati del nuovo lotto per i valori della media e della varianza calcolati in base ai lotti J del dataset Training che compongono il modello.
2. Proiettare il nuovo lotto i -esimo nello spazio del modello di monitoraggio.
3. Calcolare le statistiche T^2 e Q e confrontarle con il limite di confidenza delle carte di controllo. Se entrambi i valori sono al di sotto del limite di confidenza, si aggiorna il modello con l'inserimento del nuovo lotto e si ritorna allo step 1 altrimenti si passa allo step 4.
4. Se oltre al limite è presente un solo lotto si segnala un "Anomalia di Prodotto". Se oltre al limite è presente un gruppo di lotti si segnala un "Anomalia di Processo".

Nell'illustrazione 47 sono riportati in dettaglio i passi da compiere per l'aggiornamento aggiungendo nuovi lotti all'interno del modello.

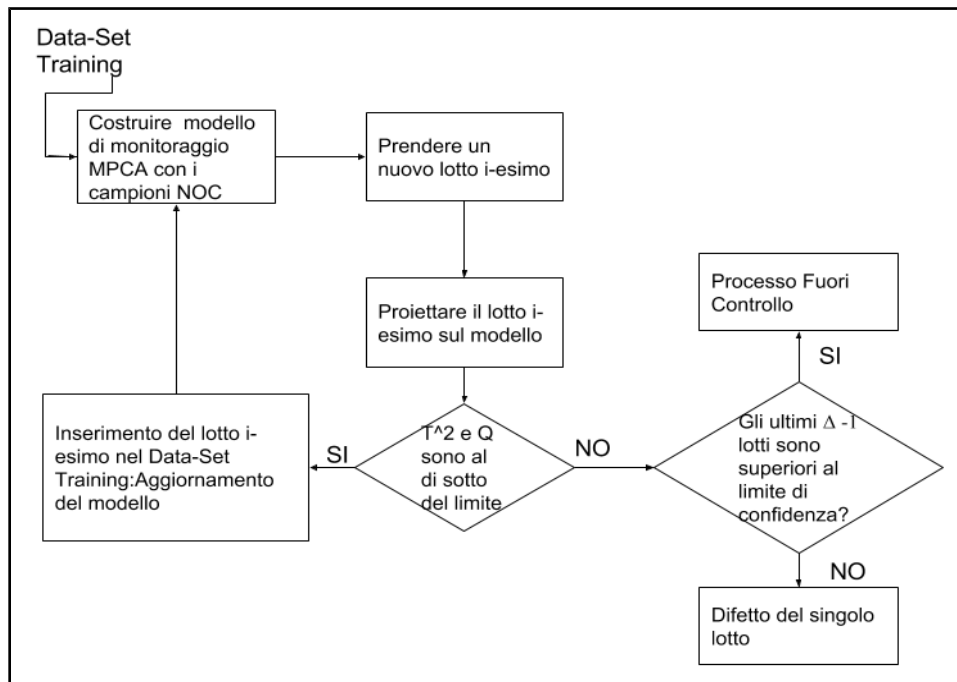


Illustrazione 47: Step di monitoraggio offline e Aggiornamento del modello

Nell'illustrazione 48, sono riportati in dettaglio le sottofasi del box “Aggiornamento del Modello” da seguire per aggiornare il modello ricevendo in input i dati dei nuovi lotti NOC.

Le sottofasi del box di Aggiornamento del Modello sono le seguenti

1. Inserimento nel dataset Training di nuovi lotti NOC;
2. Pre-processing dei dati del nuovo dataset Training;
3. Calcolo della matrice degli *Score* e dei *Loadings* del Modello MPCA;
4. Calcolo dei limiti di controllo Q e T^2 ;
5. Controllo dell'errore percentuale associato al modello. Se il modello non è migliorato si ritorna al passo 1 altrimenti viene prodotto l'output del processo di aggiornamento del modello: Modello PCA aggiornato e limiti di controllo aggiornati.

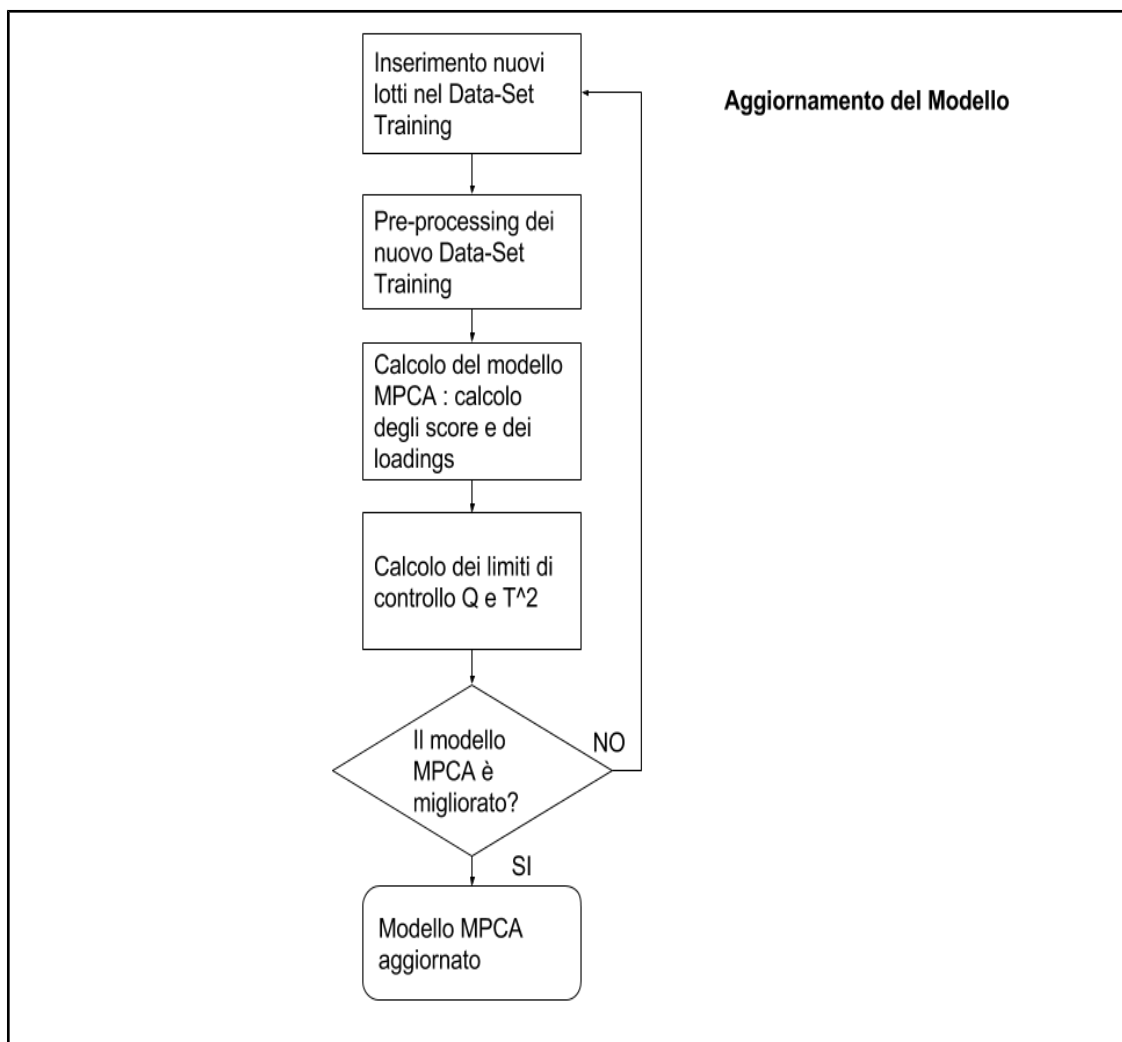


Illustrazione 48: Sottofasi del processo di Aggiornamento del modello

Conclusioni

In questa tesi è stato discusso il problema di monitoraggio di un processo batch di miscelazione finalizzato alla produzione di pellets in materiale plastico. Sono state utilizzate delle tecniche statistiche multivariate di Controllo di Processo. Sono stati elaborati i dati forniti dall'azienda Mixer Spa, relativi al processo di miscelazione, in particolare, i valori di quattro variabili di processo, misurati da sensori posti sul macchinario di miscelazione per quattro istanti di tempo successivi. Tali dati relativi a lotti produttivi, sono stati trattati con tecniche di riduzione dimensionale (MPCA e Tucker3), che hanno permesso di aggregare le informazioni sulla variabilità comune dei dati all'interno di un modello. L'utilizzo delle Carte di Controllo Multivariate, tramite delle simulazioni sui dati di processo, realizzate con il linguaggio di programmazione R, ha permesso di affrontare il problema del monitoraggio. E' stato svolto uno studio per testare la capacità delle carte di controllo costruite, di individuare lotti in condizioni anomale utilizzando due dataset di lotti storici, uno in condizioni anormali e l'altro normali. I risultati dei Test hanno dimostrato come il modello costruito sulla base del dataset *"Training"* abbia alte prestazioni sia nel riconoscere lotti in condizioni normali ed sia nel rilevare reali deviazioni dalla variabilità comune dei dati. Inoltre è stato effettuata uno studio, tramite lo strumento del *"Contribution Plot"*, volto ad individuare le variabili di processo responsabili della variabilità straordinaria all'interno di un processo. Infine è stato proposta una sequenza di Step di monitoraggio di processo, da implementare online per il controllo tra un lotto e l'altro nel processo di miscelazione; il monitoraggio online permette di intervenire tempestivamente sul prodotto ritirandolo dalla

produzione. Tale approccio di monitoraggio online, rappresenta per l'azienda fornitrice dei dati analizzati, Mixer Spa, un elemento innovativo e di miglioramento all'interno del sistema produttivo, in quanto ad oggi il controllo di qualità è effettuato unicamente offline al termine della linea produttiva. In ultimo è proposto un metodo per aggiornare il modello Multiway-PCA con l'introduzione di dati relativi a nuovi lotti. Un ulteriore sviluppo a tale sistema di monitoraggio è rappresentato dalla possibilità di associare un sistema di diagnosi delle cause delle anomalie di prodotto e di processo. Gli sviluppi futuri riguardano l'inserimento di questa tipologia di strumento all'interno di un progetto realizzato dall'azienda di consulenza Kode Srl, il cui scopo è la creazione di un sistema di controllo di processo da applicare all'intera linea produttiva dell'azienda Mixer Spa.

Appendice

Codice R

```
library("ChemometricsWithR", lib.loc=~ /R/win-library/3.1")
library("psych", lib.loc=~ /R/win-library/3.1")

#Modello
Mode<-read.table(file.choose(),header=T,sep=",")
Mode<-Mode[,-1]
Mode←as.matrix(Mode)

#Preprocessing
#preprocessig del modello:1837 lotti per 4 variabili e 4 istanti di tempo
rownames(Mode)<-1:length(Mode[,1])
Ms<-scale(Mode,center=TRUE,scale=TRUE)
Mode[1830,]

#Raw data
medie<-mean(Mode[1:1837,1])
for(i in 1:16) { if(i!=1)
  medie <-cbind(medie,mean(Mode[1:1837,i]))}
V1media<-cbind(medie[,1],medie[,5],medie[,9],medie[,13])
V2media<-cbind(medie[,2],medie[,6],medie[,10],medie[,14])
```

```
V3media<-cbind(medie[,3],medie[,7],medie[,11],medie[,15])
V4media<-cbind(medie[,4],medie[,8],medie[,12],medie[,16])
V1media<-t(V1media)
V2media<-t(V2media)
V3media<-t(V3media)
V4media<-t(V4media)
dv<-sd(Mode[1:1837,1])
for(i in 1:16) { if(i!=1)
dv <-cbind(dv,sd(Mode[1:1837,i]))
}

dv1<-cbind(dv[,1],dv[,5],dv[,9],dv[,13])
dv2<-cbind(dv[,2],dv[,6],dv[,10],dv[,14])
dv3<-cbind(dv[,3],dv[,7],dv[,11],dv[,15])
dv4<-cbind(dv[,4],dv[,8],dv[,12],dv[,16])

dv1<-t(dv1)
dv2<-t(dv2)
dv3<-t(dv3)
dv4<-t(dv4)

ts.plot(V1media,main="Variabile 1 Temperatura della mescola",xlab="istanti di
campionamento",ylim=c(700,1350),ylab="Variabile:Temperatura Mescola")
lines(V1media+2*dv1,col="red",lty=2)
lines(V1media-2*dv1,col="red",lty=2)

ts.plot(V2media,main="Variabile 2 Durata della Sotto-fase",xlab="istanti di
campionamento",ylim=c(0,50),ylab="Variabile:Durata")
lines(V2media+2*dv2,col="red",lty=2)
lines(V2media-2*dv2,col="red",lty=2)

ts.plot(V3media,main="Variabile 3 Potenza necessaria per miscelazione
",xlab="istanti di campionamento",ylab="Variabile:Energia")
```

```
lines(V3media+2*dv3,col="red",lty=2)
lines(V3media-2*dv3,col="red",lty=2)
ts.plot(V4media,main="Variabile 4 Peso della Mescola",xlab="istanti di
campionamento",ylab="Variabile:Peso Mescola")
lines(V4media+2*dv4,col="red",lty=2)
lines(V4media-2*dv4,col="red",lty=2)

#MPCA
prtrai←ChemometricsWithR::PCA(Ms)

#Modello
#numero di lotti del training
l=length(Ms)
#numero delle componenti principali
A=3
#T2
t01<-prtrai$scores[,1:3]^2%%diag((((prtrai$var[1:3]))^-1)[1:3,1:3]
A=3
T2<-rowSums(t01)
plot(T2,xlab="lotti",main="Multi-way PCA: Carta di controllo T2 del Modello")
lines(T2,xlab="lotti")
# il mio T2 ? un po piu alto
Tquadroalfa2way<-((A*(l^2-1))/(l*(l-A)))*qf(0.95,A,l-A,lower.tail = TRUE, log.p =
FALSE)
#valore limite di T^2
Tquadroalfa2way
abline(h=Tquadroalfa2way,col="red")
l<-t(prtrai$loadings[,1:3])
```

```
contributoT2<-prtrai$scores[,1:3]%*%diag(((prtrai$var[1:3]))^-1)[1:3,1:3]%*%I
fuori<-which(T2>Tquadroalfa2way)
#limite di confidenza di T2
Tquadroalfa2way
l<-t(prtrai$loadings[,1:3])
# contributo al Q da parte delle singole variabili
q.cont <- Ms- (prtrai$scores[,1:3]%*%l)
# contributio cumulativo al Q per ciasciun campione
Q <- rowSums(q.cont^2)
#calcolo i limiti di confidenza per Q
#Inverse error function
invErf <- function(x) {
  # argument x must lie between -1 and 1
  qnorm((1 + x) / 2) / sqrt(2)
}
theta1<-sum(prtrai$var[(3+1):length(prtrai$var)])
theta2<-sum(prtrai$var[(3+1):length(prtrai$var)]^2)
theta3<-sum(prtrai$var[(3+1):length(prtrai$var)]^3)
h0<-1-((2*theta1*theta3)/(3*theta2^2))
if (h0 < 0.001){
  h0 <- 0.001
}else{
  h0
}
ca <- sqrt(2)*invErf(2*0.99-1)
h1 <- (ca*sqrt(2*theta2*h0^2))/theta1
h2 <- (theta2*h0*(h0-1))/(theta1^2)
Qlim <- theta1*(1+h1+h2)^(1/h0)
```

```
plot(Q,main="Multi-way PCA: Carta di controllo Q del Modello")
abline(h=Qlim,col="red")
lines(Q)
outq<-which(Q>Qlim)

#Grafici
##Scoreplot
x <- prtrai$scores
expvar <- c(prtrai$var[1]/prtrai$totalvar*100,
            + prtrai$var[2]/prtrai$totalvar*100)
plot(x[,1], x[,2], xlab = paste('PC 1', '(exp var', round(expvar[1], 2), '%)'), main =
'Two way:Scores Plot del modello',
     ylab = paste('PC 2', '(exp var', round(expvar[2], 2), '%)'),xlim=c(-20,20),ylim=c(-
20,20))
abline(h = 0, v = 0, col = 'grey', lty = 2)

## Carta T2 e Q
plot(T2, Q, xlab = 'T2', ylab = 'Q',
     xlim=c(0,16),ylim=c(0,24),main="Carta di Controllo T2/Q")
abline(h=Qlim,lty=2)
abline(v=Tquadroalfa2way,lty=2)
out<-which(T2>Tquadroalfa2way)
#loadings plot
y<- prtrai$loadings
expvar <- c(prtrai$var[1]/prtrai$totalvar*100,
            prtrai$var[2]/prtrai$totalvar*100)
plot(y[,1], y[,2],main="Plot dei Loadings", xlab = paste('PC 1', '(exp var',
round(expvar[1], 2), '%)'),
```

```
ylab = paste('PC 2', '(exp var', round(expvar[2], 2), '%)')
text(y[1,1], y[1,2], "Temperatura1", pos=2)
text(y[2,1], y[2,2], "Durata1", pos=1)
text(y[3,1], y[3,2], "Potenza1", pos=3)
text(y[4,1], y[4,2], "Peso-Mescola1", pos=3)
text(y[5,1], y[5,2], "Temperat2", pos=1)
text(y[6,1], y[6,2], "Durata2", pos=1)
text(y[7,1], y[7,2], "Potenza2", pos=3)
text(y[8,1], y[8,2], "Peso-Mescola2", pos=1)
text(y[9,1], y[9,2], "Temperatura3", pos=1)
text(y[10,1], y[10,2], "Durata3", pos=3)
text(y[11,1], y[11,2], "Potenza3", pos=3)
text(y[12,1], y[12,2], "Peso-Mescola3", pos=4)
text(y[13,1], y[13,2], "Temperatura4", pos=3)
text(y[14,1], y[14,2], "Durata4", pos=3)
text(y[16,1], y[16,2], "Peso-Mescola4", pos=3)
abline(h=0, lty=2)
abline(v=0, lty=2)
which(y[,1]>0.2)
```

#Proiezione dei lotti test

```
NOC<-read.table(file.choose(), header=T, sep=",")
NOC<-NOC[,-1]
NOC<-as.matrix(NOC)
AOC<-read.table(file.choose(), header=T, sep=",")
AOC<-AOC[,-1]
AOC<-as.matrix(AOC)
Ms<-scale(Mode, center=TRUE, scale=TRUE)
```



```
mediecolonne<-attributes(Ms)
medcol<-mediecolonne$`scaled:center`
devcol<-mediecolonne$`scaled:scale`

#Preprocessing
#Preprocessig del test
NOCpre<-cbind((NOC[,1]-medcol[1])/devcol[1],(NOC[,2]-medcol[2])/devcol[2],
(NOC[,3]-medcol[3])/devcol[3],(NOC[,4]-medcol[4])/devcol[4],(NOC[,5]-
medcol[5])/devcol[5],(NOC[,6]-medcol[6])/devcol[6],(NOC[,7]-
medcol[7])/devcol[7],(NOC[,8]-medcol[8])/devcol[8],(NOC[,9]-
medcol[9])/devcol[9],(NOC[,10]-medcol[10])/devcol[10],(NOC[,11]-
medcol[11])/devcol[11],(NOC[,12]-medcol[12])/devcol[12],(NOC[,13]-
medcol[13])/devcol[13],(NOC[,14]-medcol[14])/devcol[14],(NOC[,15]-
medcol[15])/devcol[15],(NOC[,16]-medcol[16])/devcol[16])
AOCpre<-cbind((AOC[,1]-medcol[1])/devcol[1],(AOC[,2]-medcol[2])/devcol[2],
(AOC[,3]-medcol[3])/devcol[3],(AOC[,4]-medcol[4])/devcol[4],(AOC[,5]-
medcol[5])/devcol[5],(AOC[,6]-medcol[6])/devcol[6],(AOC[,7]-
medcol[7])/devcol[7],(AOC[,8]-medcol[8])/devcol[8],(AOC[,9]-
medcol[9])/devcol[9],(AOC[,10]-medcol[10])/devcol[10],(AOC[,11]-
medcol[11])/devcol[11],(AOC[,12]-medcol[12])/devcol[12],(AOC[,13]-
medcol[13])/devcol[13],(AOC[,14]-medcol[14])/devcol[14],(AOC[,15]-
medcol[15])/devcol[15],(AOC[,16]-medcol[16])/devcol[16])
tnew2<-(NOCpre%*%prtrai$loadings[,1:3])
test02<-tnew2[,1:3]^2%*%diag((((prtrai$var[1:3]))^1)[1:3,1:3])
T2test2<-rowSums(test02)
plot(T2test2,ylab="T2 convalida NOC",xlab="Lotti",main="Multi-way PCA: Carta
di controllo T2 lotti NOC")
lines(T2test2)
```

```
abline(h=Tquadroalfa2way,col="red")
nocT2_sopralimite<-which(T2test2>Tquadroalfa2way)
#Q e matrice dei residui
enew_test2=NOCpre-tnew2%*%I
# Q
Q_new_test2=rowSums(enew_test2^2)
plot(Q_new_test2,ylab="Q convalida NOC",main="Multiway-PCA: Carta di
controllo Q lotti NOC")
lines(Q_new_test2)
abline(h=Qlim,col="red")
nocQ_sopralimite<-which(Q_new_test2>Qlim)
residuitestaocNOC<-enew_test2^2

#Analisi delle Determinanti Contribution Plot
barplot(residuitestaocNOC[155,],,,main="Contribution plot Campione
154",cex.names=0.5,names.arg=c("Temperatura1","Tempo1","Potenza1","Peso
Mescola1","Temperatura2","Tempo2","Potenza2","PesoMescola2","Temperatur
a3","Tempo3","Potenza3","PesoMescola3","Temperatura4","Tempo4","Potenza
4","PesoMescola4"),col=c("black"))

#Analisi delle Determinanti Contribution Plot
barplot(residuitestaocNOC[156,],,,main="Contribution plot Q Campione
154",cex.names=0.5,names.arg=c("Temperatura1","Tempo1","Potenza1
1","PesoMescola1","Temperatura2","Tempo2","Potenza1
2","PesoMescola2","Temperatura3","Tempo3","Potenza1
3","PesoMescola3","Temperatura4","Tempo4","Potenza1
4","PesoMescola4"),col=c("black"))
```

#Analisi delle Determinanti Contribution Plot

```
barplot(residuitestaocNOC[157,],,main="Contribution      plot      Campione
154",cex.names=0.5,names.arg=c("Temperatura1","Tempo1","Potenza1","Peso
Mescola1","Temperatura2","Tempo2","Potenza2","PesoMescola2","Temperatur
a3","Tempo3","Potenza3","PesoMescola3","Temperatura4","Tempo4","Potenza
4","PesoMescola4"),col=c("black"))
```

#Analisi delle Determinanti Contribution Plot

```
barplot(residuitestaocNOC[158,],,main="Contribution      plot      Campione
154",cex.names=0.5,names.arg=c("Temperatura1","Tempo1","Potenza1","Peso
Mescola1","Temperatura2","Tempo2","Potenza1
2","PesoMescola2","Temperatura3","Tempo3","Potenza1
3","PesoMescola3","Temperatura4","Tempo4","Potenza1
4","PesoMescola4"),col=c("black"))
```

#Test AOC del Modello

```
tnew<-(AOCpre%*%prtrai$loadings[,1:3])
test01<-tnew[,1:3]^2%*%diag((prtrai$var[1:3])^-1)[1:3,1:3]
T2test<-rowSums(test01)
plot(T2test,ylim = c(0, max(c(T2test,Tquadroalfa2way))),ylab="T2 convalida
AOC",xlab="Lotti",main="Multi-way PCA: Carta di controllo T2 lotti AOC")
lines(T2test)
abline(h=Tquadroalfa2way,col="red")
T2aocsopralimite<-which(T2test<Tquadroalfa2way)
#Control Chart Q
enew_test=AOCpre-tnew%*%t(prtrai$loadings[,1:3])
```

```
Q_new_test=rowSums(enew_test^2)
plot(Q_new_test,xlab="lotti",ylab="Q convalida AOC",main="Multi-way PCA:
Carta di controllo Q lotti AOC")
abline(h=Qlim,col="red")
which(Q_new_test>10000)
lines(Q_new_test)
residuitestaocAOC<-enew_test^2
```

#Analisi delle Determinanti Contribution Plot

```
barplot(residuitestaocAOC[56,],,main="Contribution plot Campione
154",cex.names=0.5,names.arg=c("Temperatura1","Tempo1","Potenza1","Peso
Mescola1","Temperatura2","Tempo2","Potenza1
2","PesoMescola2","Temperatura3","Tempo3","Potenza1
3","PesoMescola3","Temperatura4","Tempo4","Potenza1
4","PesoMescola4"),col=c("black"))
```

PLOT Carta T2 e Q

```
plot(T2, Q, xlab = 'T2', ylab = 'Q', main = 'T2/Q residuals',
xlim=c(0,50),ylim=c(0,50))
abline(h=Qlim,lty=2)
abline(v=Tquadroalfa2way,lty=2)
points(T2test,Q_new_test,col="red")
points(T2test2[46:146],Q_new_test2[46:146],col="green")
```

##Scoreplot

```
x <- prtrai$scores
expvar <- c(prtrai$var[1]/prtrai$totalvar*100,
+ prtrai$var[2]/prtrai$totalvar*100)
+plot(x[,1], x[,2], xlab = paste('PC 1', '(exp var', round(expvar[1], 2), '%)'), main =
```

```
'Two way:Scores Plot del modello e dei lotti test NOC e AOC', ylab = paste('PC 2',  
'(exp var', round(expvar[2], 2),'%)'),xlim=c(-20,20),ylim=c(-20,20))  
points(tnew2,col="green")  
points(test01,col="red")  
abline(h =0, v = 0, col = 'grey', lty = 2)
```

Bibliografia

- Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., & Yener, B. (2007). Multiway analysis of epilepsy tensors. *Bioinformatics*, 23(13), i10–i18. <http://doi.org/10.1093/bioinformatics/btm210>
- Analysis, P. C. (n.d.). Lecture 15 : Principal Component Analysis. *Intelligent Data Analysis*, 1–6.
- Auer, P. (2007). Choosing principal components : a new graphical method based on Bayesian model selection.
- Booksh, K. S., Bronk, B., & Czege, J. (2009). 3.09 - Three-Way Calibration. *Comprehensive Chemometrics*, 379–412. <http://doi.org/10.1016/B978-044452701-1.00078-8>
- Camacho, J. (2007). New Methods Based on the Projection to Latent Structures for Monitoring, Prediction and Optimization of Batch Processes. *Upv*.
- Garcia, M., Ruiz, M., Colomer, J., & Melendez, J. (2015). Multiway principal component analysis and case base reasoning methodology for abnormal situation detection in a nutrient removing SBR. *2007 European Control Conference, ECC 2007*, 5354–5360. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-84927731903&partnerID=40&md5=66bbb4e415d71fd4bea3be7231cfc818>
- Gunther, J. C., Conner, J. S., & Seborg, D. E. (2007). Fault detection and diagnosis in an industrial fed-batch cell culture process. *Biotechnology Progress*, 23(4), 851–7. <http://doi.org/10.1021/bp070063m>
- Haridy, S., & Wu, Z. (2009). Univariate and multivariate control charts for

- monitoring dynamic-behavior processes: a case study. *Journal of Industrial Engineering and Management*, 2(3), 464–498.
<http://doi.org/10.3926/jiem.2009.v2n3.p464-498>
- Hu, K., & Yuan, J. (2008). Multivariate statistical process control based on multiway locality preserving projections. *Journal of Process Control*, 18, 797–807. <http://doi.org/10.1016/j.jprocont.2007.11.002>
- Kassidas, A., Macgregor, J. F., & Taylor, P. a. (1998). Synchronization of batch trajectories using dynamic time warping. *AIChE Journal*, 44(4), 864–875.
<http://doi.org/10.1002/aic.690440412>
- Kourti, T. (2005). Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing*, 19(4), 213–246.
<http://doi.org/10.1002/acs.859>
- Laursen, K., Frederiksen, S. S., Leuenhagen, C., & Bro, R. (2010). Chemometric quality control of chromatographic purity. *Journal of Chromatography A*, 1217(42), 6503–10. <http://doi.org/10.1016/j.chroma.2010.08.040>
- MacGregor, J. F., & Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403–414.
[http://doi.org/10.1016/0967-0661\(95\)00014-L](http://doi.org/10.1016/0967-0661(95)00014-L)
- Nomikos, P., & MacGregor, J. F. (1995). Multivariate Statistical Process Control Charts for Monitoring Batch Processes.Pdf. *Technometrics*.
[http://doi.org/doi:10.1016/0967-0661\(95\)00014-L](http://doi.org/doi:10.1016/0967-0661(95)00014-L)
- Reinikainen, S.-P., & Höskuldsson, A. (2007). Multivariate statistical analysis of a multi-step industrial processes. *Analytica Chimica Acta*, 595(1-2), 248–56. <http://doi.org/10.1016/j.aca.2007.02.019>
- Ruiz, M., Colomer, J., & Mel, J. (n.d.). Multiway Principal Component Analysis and Case Base-Reasoning approach to situation assessment in a WasteWater Treatment Plant.
- Shen, F., Ge, Z., & Song, Z. (2015). Multivariate Trajectory-Based Local Monitoring Method for Multiphase Batch Processes. *Industrial & Engineering Chemistry Research*, 54(4), 1313–1325.

<http://doi.org/10.1021/ie503921t>

Van sprang, E. N. M., Ramaker, H. -J, Westerhuis, J. A., Smilde, A. K., & Wienke, D. (2005). Statistical batch process monitoring using gray models. *AIChE Journal*, 51(3), 931–945.

<http://doi.org/10.1002/aic.10348>

Westerhuis, J. a., Gurden, S. P., & Smilde, A. K. (2000). Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 51(1), 95–114.

[http://doi.org/10.1016/S0169-7439\(00\)00062-9](http://doi.org/10.1016/S0169-7439(00)00062-9)

Wise, B. M., Gallagher, N. B., Butler, S. W., White, D., & Barna, G. G. (1997). Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: impact of measurement selection and data treatment on sensitivity. *Ifac Safeprocess*, 97, 35–42.

Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.25.7702&rep=rep1&type=pdf>

Zhao, C., Wang, F., Mao, Z., Lu, N., & Jia, M. (2008). Improved Batch Process Monitoring and Quality Prediction Based on Multiphase Statistical Analysis. *Industrial & Engineering Chemistry Research*, 47(3), 835–849. <http://doi.org/10.1021/ie0707624>

Ringraziamenti

Desidero ringraziare innanzitutto il mio relatore di Tesi, il Professore Riccardo Dulmin, per avermi seguito durante il mio percorso di Tesi.

Ringrazio la società Kode srl che ha permesso lo svolgimento di un tirocinio della durata di 6 mesi, durante il quale ho avuto la possibilità di conoscere e avvicinarmi al campo del Data Science.

Ringrazio, in particolare, il mio tutor aziendale e relatore di Tesi, Marco Calderisi, per avermi seguito nei mesi di tirocinio e nella scrittura della Tesi. Questo periodo, mi ha permesso di apprendere sia nozioni in campo statistico e informatico, grazie all'utilizzo del linguaggio di programmazione R. Vorrei inoltre ringraziare il mio tutor per la sua disponibilità e la sua collaboratrice Ilaria Ceppa, che mi ha aiutato nel colmare le mie carenze informatiche, entrambi sono riusciti a trasmettermi una grande passione per il loro lavoro.

Ringrazio le amiche e colleghe musiciste, in particolare Alessandra e Aurora, che da circa un anno e mezzo condividono con me, ogni avvenimento della giornata, e soprattutto la passione per la musica all'interno dell'Orchestra dell'Università di Pisa.

Ringrazio i miei colleghi di università, con cui ho affrontato moltissimi ostacoli, che mi hanno accompagnato nel mio percorso di Laurea. Ringrazio in particolare Erika perchè è stata la prima persona che ho conosciuto del mio corso, ringrazio Valeria che mi ha fatto conoscere la bellissima Castellina Marittima e Michela che mi ha sopportato in questo ultimo mese in interminabili chiamate.

Come non ringraziare i miei coinquilini storici Fefè, Angelo e Annamaria, è difficile trovare le parole per definire cosa sono stati questi anni, dire che abbiamo condiviso un pezzo di vita insieme è poco, siamo cresciuti e superato mille ostacoli/esami insieme grazie alla vicinanza degli altri, probabilmente mi azzardo che a dire che abbiamo condiviso insieme gli anni più belli e per questo non vi dimenticherò mai e ringrazio anche il mio nuovo coinquilino Dario che è bravissimo cuoco .

In ultimo ringrazio i miei amici di sempre, con cui sono cresciuta nel mio paese, che ci saranno sempre per me, come io per loro, ed infine ringrazio i miei genitori e i miei nonni che mi hanno permesso di affrontare questo percorso, visto che con le parole non sono brava, colgo l'occasione per dirgli quanto gli voglio bene.